DOCUMENT RESUME

ED 195 573 TM 810 023

AUTHOR Lohman, David F.
TITLE Spatial Ability: Individual Differences in Speed and Level. Technical Report No. 9. Aptitude Research Project.
INSTITUTION Stanford Univ., Calif. School of Education.
SPONS AGENCY Advanced Research Projects Agency (DOD), Washington, D.C.: Office of Naval Research, Arlington, Va. Personnel and Training Research Programs Office.
PUB DATE Oct 79
CONTRACT N00014-75-C-0882
NOTE 163p.

EDRS PRICE MF01/PC07 Plus Postage.
DESCRIPTORS Cognitive Measurement: *Cognitive Processes: *Conceptual Tempo: *Difficulty Level: Higher Education: *Individual Differences: Males: Secondary Education: Spatial Perception: *Verbal Ability
IDENTIFIERS *Spatial Ability: Spatial Tests

ABSTRACT

Previous research on the nature of aptitude processes had ignored the important differences between speed and level of performance. However, investigations of the relationship between individual differences in speed and level suggest that speed of solving simple spatial tasks is largely independent of level attained on more complex spatial tasks. The present experiment was designed to investigate the relationships between speed, level, and complexity in individual differences in spatial ability. Thirty male high school and college students were selected to represent a wide range of individual differences in verbal ability and spatial ability. Subjects were required to match, rotate, combine, and combine and then rotate three to eight point polygons. Each subject was administered 216 items. Data were analyzed using regression, analysis of variance, and correlational techniques. Correlations between intercepts, slopes, predicted times at maximum difficulty, and level scores and reference constructs supported the basic hypothesis of speed-level independence. In particular, individual differences in speed of solving simple items were independent of both level on the task and reference abilities. Speed of solving complex items was strongly correlated with level on the task, but more highly correlated with verbal than spatial reference constructs. (Author/RL)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED195573

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

SPATIAL ABILITY: INDIVIDUAL DIFFERENCES IN SPEED AND LEVEL

DAVID F. LOHMAN

TECHNICAL REPORT NO. 9
APTITUDE RESEARCH PROJECT
SCHOOL OF EDUCATION
STANFORD UNIVERSITY

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
Office of Naval Research
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

Sponsored by

Personnel and Training Research Programs
Psychological Sciences Division
Office of Naval Research

and

Advanced Research Projects Agency

under

Contract No. N00014-75-C-0882

Approved for public release; distribution unlimited.
Reproduction in whole or in part is permitted for any purpose of the United States Government.

TM 810 023

OCTOBER 1979

SPATIAL ABILITY: INDIVIDUAL DIFFERENCES IN SPEED AND LEVEL

DAVID F. LOHMAN

TECHNICAL REPORT NO. 9
APTITUDE RESEARCH PROJECT
SCHOOL OF EDUCATION
STANFORD UNIVERSITY

Sponsored by

Personnel and Training Research Programs
Psychological Sciences Division
Office of Naval Research

and
Advanced Research Projects Agency

under

Contract No. N00014-75-C-0882

The views and conclusions contained in this document are those of the author and should not be interpreted as necessarily representing the official policies, either expressed or implied, of the Office of Naval Research, the Advanced Research Projects Office, or the U.S. Government

Approved for public release; distribution unlimited. Reproduction in whole or in part is permitted for any purpose of the United States Government.

OCTOBER 1979

DEC 1 9 1980

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER 9	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) SPATIAL ABILITY: Individual Differences in Speed and Level		5. TYPE OF REPORT & PERIOD COVERED Technical Report #9 6. PERFORMING ORG. REPORT NUMBER 9
7. AUTHOR(s) David F. Lohman		8. CONTRACT OR GRANT NUMBER(s) N00014-75-C-0882
9. PERFORMING ORGANIZATION NAME AND ADDRESS School of Education Stanford University Stanford, California 94305		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS NR 154-376
11. CONTROLLING OFFICE NAME AND ADDRESS Personnel and Training Research Programs Psychological Sciences Division, ONR, 458		12. REPORT DATE October 1979 13. NUMBER OF PAGES 133
14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office)		15. SECURITY CLASS. (of this report) UNCLASSIFIED 15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

UNLIMITED

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

UNLIMITED

18. SUPPLEMENTARY NOTES

This research was jointly sponsored by the Office of Naval Research and the Defense Advanced Research Projects Agency.

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Spatial ability, mental rotation, mental construction, individual differences in cognitive processes, task complexity, individual differences in speed, individual differences in level, aptitude processes.

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

This experiment investigated the relationships between speed, level, and complexity in individual differences in spatial ability. Thirty male high school and college students representing a wide range of individual differences in verbal ability and spatial ability were required to match, rotate, combine, and combine and then rotate three to eight point polygons. Three levels of construction (one, two, or three stimuli), three levels of stimulus complexity (low, medium or high), three levels of rotation

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE
S/N 0102-LF-014-6601

UNCLASSIFIED
SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

(0, 90, or 180 degrees), and two types of discriminative response (correct or incorrect) were fully crossed in the within subjects design. Additionally, type of addition (left or right) and complexity of the to-be-constructed stimulus (low or high) were nested within levels of the construction facet but crossed with each other and all other facets.

Correctness, confidence, total time, and from two to four component times were obtained on each item. Data were analyzed using regression, analysis of variance, and correlational techniques.

In the regression analyses, latency measures were regressed on percent of the sample failing each item. Correlations between intercepts, slopes, predicted times at maximum difficulty, and level scores and reference constructs supported the basic hypothesis of speed-level independence, In particular, speed of solving simple items was independent of both level on the task and reference abilities. Speed of solving complex items was strongly correlated with level on the task, but more highly correlated with verbal than spatial reference tests.

The major results of the means analyses were: a) there were large facet interactions in all analyses, implying that complexity increased nonlinearly; b) encoding time increased linearly with increases in stimulus complexity; c) rotation time and errors increased almost linearly over the three levels of rotation; when errors and latency were combined, the increase was perfectly linear; d) rotation was unaffected by stimulus complexity, but slower for constructed than nonconstructed stimuli; e) construction was differentially affected by the number of stimuli to be combined, the complexity of these stimuli, the location of the addition, and, most significantly, the complexity of the to-be-constructed stimulus; f) for match times "yes" responses were faster on correct items and "no" responses faster on incorrect items. Those who performed well on the experimental task showed little increase in match time over the rotation and construction facets, while those who performed poorly showed a linear increase in match time over the rotation facet and a large increase on construction items.

Relationships between individual differences in total scores, cell totals, and the pattern of cell means over design facets were explored in the correlational analyses. In general, correlations between correctness indices and reference tests supported the construct validity of the task and its facets, while correlations between latency indices and reference tests suggested more about the processes or strategies subjects were using to solve items. The major results were: a) total number correct was strongly correlated with reference spatial tests, while average total time was weakly related to verbal reference tests; b) steep slopes were related to verbal ability for the regressions of encoding time on stimulus complexity; rotation time on rotation and stimulus complexity; and construction time on construction, stimulus complexity, and type of addition; c) a linear decline in correctness over the rotation facet was related to high scores on two speeded rotation reference tests (Cards and Figures), while a linear decline in correctness over the construction facet was related to high scores on two construction reference tests (Form Board and Object Assembly).

It was concluded that the results of the experiment supported the hypothesis of speed-level independence and provided important insights into the nature of spatial ability.

PREFACE

The investigation reported herein is part of an ongoing research project aimed at understanding the nature and importance of individual differences in aptitude for learning. Requests for information regarding this project and for copies of this or other technical reports should be addresses to:

Professor Richard E. Snow, Principal Investigator
Aptitude Research Project
School of Education
Stanford University
Stanford, California 94305

TABLE OF CONTENTS

PREFACE ii

LIST OF TABLES vi

LIST OF FIGURES vii

Chapter

I INTRODUCTION 1

Previous Approaches to the Problem 1

Selection of Experimental Task 2

The Speed-Level Problem 3

Limitations of Previous Methods 4

Factor Analytic Evidence 4

Speed-Level Studies 5

Complexity 6

Unpacking Difficulty 8

Varieties of Individual Differences in Spatial Ability 11

Correlational Studies 11

Experimental Studies of Spatial Thinking 14

Individual Differences in the Experimental Studies 15

Solution Strategies 16

Methodological Issues 17

Limitations of Tests and Factors 17

Contrast Correlations 19

Conclusions 19

Overview of the Study 20

II DESIGN AND PROCEDURE 21

Design 21

Experimental Task 21

Draw Trials 23

Simple Reaction Time Control 23

Reference Constructs 23

Assignment of Items to Blocks 24

Subjects 24

Stimuli 28
Overview of Trial Types 28
Construction of Stimuli 28
Construction of Construction Items 30
Apparatus 32
Procedure 32

III RESULTS AND DISCUSSION 34

Overview 34
I. Preliminary Analyses 35
Dependent Measures 35
Outliers 35
A More Stringent Definition of Correctness 36
II. Regression Analyses 36
Construct Validity of the Level Scores 38
Speed-Level Correlations 40
III. Separate Analyses on Total Time, Correctness, and Confidence 44
Practice Effects 44
RT for Incorrect Items 44
Analysis on Total Time 45
Means Analysis 45
Objections to Contrast Correlations 49
Total Time Contrast Correlations 53
Analyses on Correctness 57
Means Analysis 57
Contrast Correlations 60
Preliminary Analysis of Confidence 60
Comparison of SCORE3 and Total Time Analyses 62
Summary of the Separate Analyses 64
IV. Combined Analyses on Total Time and Correctness 65
Sum and Difference Score Analyses 67
Interpretation of the Sum and Difference Score Analyses . . 68
Defense of Sum and Difference Scores 76
Summary of the Sum and Difference Score Analyses 78

V. Component Time Analyses 79
Encoding 80
Means Analysis 80
Contrast Correlations 80
The Construction Component 82
The Rotation Component 90
No Construction Items 91
Correlations 91
Rotation of Constructed Stimuli 98
Match Component Times 98
All Match Times 98
Correlations 100
Summary of Match Component Analyses 103
VI. Comparison of Overall and Component Analyses 105
Total Scores 105
Encoding 106
Construction 106
Rotation 109
Match 110
VII. General Discussion 111
IV SUMMARY 116
Total Scores 117
Encoding 117
Construction 118
Rotation 121
Match 122
V CONCLUSIONS 124
REFERENCES 127

LIST OF TABLES

Table

3.1 Correlations between Total SCORE, Total SCORE3, Total Correct Draw Trials, and Reference Tests (N=30) 39

3.2 Correlations between Slopes, Intercepts, Predicted Times at Maximum Difficulty, and Reference Tests (N=30) 41

3.3 Correlations between Slopes, Intercepts, Predicted Times at Maximum Difficulty, and Reference Tests (N=29) 42

3.4 Analysis of Variance on Total Time 46

3.5 Correlations between Total Time Main Effect Contrast Scores and Reference Tests (N=30) 54

3.6 SCORE Analysis of Variance 58

3.7 SCORE3 Analysis of Variance 59

3.8 Correlations between SCORE3 Main Effect Contrast Scores and Reference Tests (N=30) 61

3.9 ZSUM Analysis of Variance 69

3.10 ZDIFF Analysis of Variance 70

3.11 Analysis of Variance for Encoding Component Time 81

3.12 CSUM Analysis of Variance 84

3.13 CDIFF Analysis of Variance 85

3.14 Correlations between Construction CSUM, Component Time, and SCORE3 N Contrasts and Reference Tests (N=30) 89

3.15 Analysis of Variance for Rotation Component Time 92

3.16 Analysis of Variance for Rotation SCORE3 93

3.17 Correlations between Mean SCORE, Mean Rotation Component Time and Reference Tests, by Block (N=30) 97

3.18 Correlations between Mean Match Time and Reference Tests . . 104

LIST OF FIGURES

Figure

2.1 Design of the experimental task 22

2.2 Scatterplot of scores on the two spatial composites for all 158 males in the reference population 26

2.3 Scatterplot of scores on the first spatial and verbal composites for all 158 males in the reference population . . 27

2.4 Examples of the major trial types in the experimental task . . 29

2.5 Examples of stimuli at each of the three levels of stimulus complexity 31

3.1 Distribution of match component times and the location of TMATCHN for subject HF 37

3.2 Plots of the eight largest two way interactions for total time 50

3.3 Plots of the rotation by product complexity interaction for total time, SCORE, SCORE3, ZSUM, and ZDIFF 66

3.4 Partitioning of variance for ZDIFF and ZSUM 71

3.5 Plots of main effect cell means for total time, SCORE3, ZDIFF, and ZSUM 73

3.6 Plots of (a) rotation component time, and (b) SCORE3, by block for the 48 no construction items 94

3.7 Plots of the rotation by stimulus complexity interaction, by block 95

3.8 Plot of rotation component time by construction 99

3.9 Mean match time for correct and incorrect items, by group . . 101

3.10 Match time for correct and incorrect items for (1) low, (2) medium, and (3) high performing groups, over (a) match, (b) rotation, (c) construction, and (d) product complexity . . . 102

CHAPTER I
INTRODUCTION

Research on aptitude for learning has entered a new era. Instructional studies have established that individual differences among learners often interact with instructional treatment variables (Cronbach and Snow, 1977; Snow, 1977). Much of this work has also underscored the need for deeper, more process oriented understanding of the psychological nature of aptitudes. Cognitive psychologists have begun the experimental analysis of individual differences in information processing, and there is now reason to hope that coordination of these lines of work will lead to process theories of aptitude for learning from instruction (Snow, 1976).

Previous Approaches to the Problem.

Three paradigms have dominated previous attempts to bridge the gap between trait and process descriptions of individual differences in aptitude.

1. Direct modeling of the cognitive task, usually with latency as the dependent measure. Much of modern cognitive psychology could be classified here. Particularly relevant examples are the work of Clark and his associates on verbal comprehension (Clark and Chase, 1972), and Shepard and his associates on mental rotation (Shepard and Metzler, 1971; Shepard and Feng, 1972; Cooper and Shepard, 1976). Although the bulk of this work ignores individual differences in performance, such need not be the case. The most noteworthy example is the extensive work of Sternberg (1977) on analogical reasoning. In addition to examining a number of sources of individual differences in cognitive performance, Sternberg also provides for the possibility of modeling error as well as latency data.

2. Computer Simulation. This approach is exemplified in the work of Simon and Kotovsky (1963) and Kotovsky and Simon (1973) on letter series items. In simplest terms, the goal of this approach is to develop computer programs that solve various tasks with the same relative ease that human subjects solve similar items. The program is considered a good model of human behavior if item solution times parallel those for human subjects.

3. Correlating information processing parameters with aptitude constructs. This approach is exemplified in the work of Hunt, Frost, and Lunneborg (1973) on the relationships between verbal and quantitative abilities and various task parameters. Similar work with a wider range of aptitude constructs has been reported by Snow, Marshalek, and Lohman (1976). Although there are a number of variants of this technique, the goal is to find a set of process parameters that, collectively, predict an aptitude construct; or, alternately, a set of aptitude constructs that predict a process parameter.

Selection of Experimental Task.

Most investigations of individual differences in cognitive processes have attempted to model performance on a task similar to an established test (e.g., Sternberg, 1977; Simon and Kotovsky, 1963). Since individual differences on the source test usually have a known network of relations with other mental tests, the investigator ordinarily does not have to establish the construct validity of his task. The main disadvantages of this approach are: a) the sources of inter-item variation must first be "discovered," b) once enumerated, they must be experimentally disentangled, and c) these experimental manipulations may destroy the construct validity of the test, which was the main reason for

using it in the first place. Since the psychometrically sound test may be a veritable hodge-podge of test facets, uncovering all the dimensions of inter-item variation is frequently a difficult endeavor. Further, once identified, it may be impossible to separate these dimensions experimentally. For example, in Paper Folding (French, Ekstrom, and Price, 1963), number of folds is necessarily confounded with number of holes in the final unfolded stimulus.

An alternative approach is to construct a new task in which a limited number of design facets are varied systematically and, if possible, orthogonally. The disadvantage of this method is that the investigator must then establish the construct validity of the task. There is no known network of relationships with other tests to fall back on.

The Speed-Level Problem

Limitations of Previous Methods.

The most glaring deficit of previous research in this area has been the unwitting assumption that speed of performance is psychologically the same as power or level of performance.

Although the distinction between speed and power is not synonymous with the difference between latency and correctness, those who model cognitive performance using reaction time data are usually studying speed of performance. In any experiment of this sort, it is necessary to "keep the error rate down" in order to minimize missing data. This is usually done by keeping items simple, thus eliminating individual differences in power. Ignoring power seems to have been a concession to methodological and statistical limitations, not a deliberate choice.

Focusing on speed of solving relatively simple items would not be a serious limitation if speed and power reflected the same psychological functions. However examination of the literature on the relationship between speed and power suggests that, particularly in the area of spatial thinking, speed of solving simple items is largely independent of power in solving more complex items of the same type (Lohman, 1979a).

Perhaps the most serious limitation in modeling speed rather than level of cognitive performance arises from the fact that power in one domain correlates strongly with power in other cognitive domains, and, more importantly, with external criteria such as learning in academic situations. On the other hand, speed of solving a simple task of one type is largely independent of speed of solving other simple tasks. Further, these various indices of speed are only slightly correlated with power in their respective domains, or to external variables such as academic achievement (Lohman, 1979a).

Factor Analytic Evidence.

This relationship between speed, power, and complexity repeatedly surfaced in a recent reanalysis of the major American factor analytic research on spatial ability (Lohman, 1979a). The studies were reanalyzed using hierarchical factor methods and nonmetric multidimensional scaling. Simple, highly speeded tests invariably defined specific factors which fell in the lower stratum of a hierarchical model or at the periphery of a two or three dimensional nonmetric scaling representation. Further, these sorts of factors emerged only when extremely similar tests were included in the analysis. On the other hand, power tests and their factors invariably fell in the upper levels of the hierarchical model or near the center of the scaling representations. Tests that defined these power factors were complex, yet often dissimilar in other respects.

The most consistent factorial split was between complex, power tests (e.g., Paper Folding and Surface Development) that defined a factor called "Visualization" and simple, highly speeded tests (e.g., Cards and Flags) that defined the "Spatial Relations" factor.

Finally, an important conclusion of this review was that the broad power factors cannot be subdivided into the various primaries of the Thurstone (1938) or French (1951; French, Ekstrom and Price, 1963) variety. Alternately, second order factors that are extracted from a matrix of primary factor

correlations do not coincide with corresponding power factors. This nonhierarchical nature of the factors most likely reflects the independence of speed and power in cognitive performance.

An important implication of this is that studying speed of solving simple tasks sheds little light on the nature of higher order aptitude constructs such as "intelligence," "verbal ability" or "spatial ability." This is especially troublesome for educational applications, since it is these more general, power aptitudes which will most likely interact with educational treatments.

The data of Hunt et al., (1973) and Snow et al., (1976) support this argument. Both of these investigations found that parameters derived from various cognitive tasks were only weakly related to higher order aptitude constructs. While some of these correlations were statistically significant, none were very large.

Speed-Level Studies.

A review of the literature on the relationship between speed and power (Lohman, 1979a) indicated that the relationship is probably moderated by the following factors:

1. Difficulty. Speed and power probably correlate differently when speed is measured over simple tasks than when it is measured over complex tasks.
2. Content Area. The relationship is probably different within verbal tests, within spatial tests, etc.
3. Accuracy. Correct responses are generally faster than incorrect responses for moderately difficult to difficult items. On simple items, however, incorrect responses are usually faster.
4. Correct "Yes" vs. Correct "No." Correct "yes" responses are usually faster than correct "no" responses, although some subjects evidence the opposite pattern on some spatial tasks (Cooper, 1976). Further, latency for these two types of

correct responses may relate differently to difficulty, and thus to complexity.

5. Guessing. The error variance introduced by this factor can seriously cloud the relationship between speed and power. Further, there are differences in willingness to guess, both between subjects and within subjects across tasks and situations. The effect of guessing is most pronounced in experiments where a yes/no response is required.
6. Alternative Solution Strategies. Many tasks can be solved in more that one way. There is some evidence that particular tests (such as the Guilford-Zimmerman Spatial Orientation test) are especially vulnerable. Items where several multiple choice alternatives are provided are also particularly suspect.
7. Motivation. Thorndike (1926), Thurstone (1937) and Furneaux (1961) all agree that motivation (or persistence) can influence both level and speed. Further, the relationships between motivation, speed and power are probably not linear; one can literally "try too hard."

Complexity

It was noted that the speed-power difference in spatial factors is confounded with differences in test complexity. Speed is measured when test items are simple, power when items are complex. But what is complexity?

It is important to distinguish between stimulus complexity and processing complexity. Stimulus complexity has been defined in a number of ways. Garner found that judged goodness of matrix patterns was highly correlated with equivalence set size (Garner, 1974; Garner and Clement, 1963). Attneave (1957), Arnoult (1960) and Stenson (1966) found that complexity ratings of random forms were strongly correlated with the number of turns or points in the form, the ratio of the perimeter squared to the area of the

form (compactness), and the variability of the angles on the perimeter of the form.

Stimulus complexity may moderate the relationship between speed and level only if increases in stimulus complexity affect corresponding increases in processing complexity. Cooper (1975) and Cooper and Podgorny (1976) claim that stimulus complexity does not affect the rate of mental rotation. However, they employed only well learned stimuli in their experiments. The stimulus complexity of unfamiliar stimuli may alter the rate of mental rotation. This would help explain why capital letters can be rotated at a much greater rate than other stimuli (see Cooper and Shepard, 1976, for other explanations of this phenomenon). Finally, an important but overlooked possibility is that the crucial aspect of stimulus complexity in many difficult spatial tests is not the complexity of the stimulus figure that is given, but rather the complexity of the figure that must be mentally constructed (c.f., Shepard and Feng, 1972).

Processing complexity, on the other hand, has received relatively little attention. It is most likely a function of the nature, number, and diversity of processes executed during the solution of a task. Certain processes may be more complex than others. Executing a process several times may be more difficult than executing it once. Similarly, switching between two different processes may be more difficult than executing the same process twice. Thus, the more complex spatial tests usually require the subject to employ several different processes, such as mental folding plus mental construction. Alternately, they may require the repeated application of one process, such as mental rotation (e.g., as in the Guilford-Zimmerman Spatial Visualization test). Finally, they may require the execution of a relatively simple set of processes on a complex, unfamiliar stimulus (e.g., as in Graham and Kendall's (1948) Memory for Designs).

Unpacking Difficulty.

The key to the resolution of the speed-power problem is in the construction of psychologically meaningful indices of item difficulty. This is perhaps best accomplished by equating difficulty with complexity of processing requirements and then constructing information processing models that account for the various dimensions of task complexity.

A task like Block Rotation (Shepard and Metzler, 1971) provides a good example of how the term "item difficulty" can be unpacked in a psychologically meaningful way. The usual procedure for determining item difficulty is to compute the proportion of the sample passing (or failing) each item. However, since only a handful of subjects pass the most difficult items, the correlation between speed and power would be based on the entire sample for the easiest item and on no subjects for the most difficult item. Another limitation is that these group statistics may not apply at the individual level. What is difficult for one subject may be relatively easy for another.

At the individual level, item difficulty should be related to both correctness and latency. Those items that the subject solves correctly are obviously easier for him than those he fails. Within the set of correctly answered items, it is likely that it will take him longer to solve the items that he finds more difficult. However, difficulty is then defined in terms of latency, and so the two cannot be separated.

Even if the group difficulty scale were a reasonable approximation for most individuals, one would still be faced with the problem of determining why some items were more difficult than others. If inter-item differences in processing complexity were unidimensional (e.g., items differ only in the degree of angular separation between the two stimulus figures) then there is a good chance that one could "discover" this dimension by examining the group difficulty scale. However, if the inter-item complexity is two, three or four dimensional, there is little chance that one could intuit these dimensions from an inspection of the group difficulty scale. This becomes obvious when one

realizes that there may be interactions between the different complexity facets, and interactions between subjects and the various complexity facets. Further, it is highly unlikely that all facets of complexity present in the test will be represented in an orderly (factorial) fashion. The psychometrically sound test may be a veritable hodge-podge of facets.

Therefore, rather than attempting to analyze a complicated existing test, it would appear more reasonable to start from the other end, by constructing a multi-faceted test. For example, one could vary the dimensionality of the stimulus figures (two versus three); the number of rotations required, in addition to the angular value of each rotation (as in the Guilford-Zimmerman Spatial Visualization test); and the like.

Returning to the block rotation task, assume that item complexity is entirely a function of the angular separation between the two stimulus figures. If item difficulty is captured by this unidimensional scale, then the following conditions should hold:

1. For the group data, both mean latency for each item type and proportion of the total group failing each item type should be monotonically increasing functions of complexity (i.e., angular separation).
2. For the individual data, average (within subject) latency of correct responses for each item type should be a monotonically increasing function of complexity. Further, individuals should not skip points on the scale. Thus, within the limits of measurement error, if an individual correctly solves item type k he must also have solved all previous item types 1 to k-1.

Assuming that the regression of log time on angular separation is linear, it is then possible to compute these regressions for the group and individual data. Both the slopes and intercepts should be positive for the group data as well as for each subject.

If either parameter is significantly negative for any

subject, it suggests that the information processing model is invalid for that subject. Further, the intercept must be significantly different from zero, while the slope may be zero. In the case of the intercept, this is because processing even the simplest item must take some time. On the other hand, the slope may be zero if it takes a subject as long to process a complex item as it does to process a relatively simple item. However, this also implies that the facet in question is not a meaningful aspect of complexity for that subject.

The correlation between the intercepts and the lengths of the regression lines yields the correlation between speed of performing the easiest item types and the level reached on more difficult item types. The individual regression lines can then be projected so that they all extend to the point of maximum complexity. At this point the correlation between these (for the most part, predicted) latencies and the known lengths of the regression lines yields the best estimate of the relationship between speed of performing complex tasks and level of performance on those tasks. Projected or known latencies at intermediate points on the scale can also be correlated with level to yield intermediate values of the speed-level correlation.

When formulated in this manner, it is obvious that the correlation between speed and power will remain constant over the range of item complexity only if there are no individual differences in the regression slopes. This has important implications both for the speed-power problem and for the generalizability of information processing parameters derived from simple tasks. If individual regression slopes are parallel, then the relationship between speed and power is constant throughout the range of complexity represented in the analysis. Parallel regression slopes also imply that individual differences in speed of solving simple tasks generalizes to the speed of solving complex tasks of the same type.

On the other hand, if there are individual differences in slopes, then speed of solving simple tasks does not generalize to

speed of solving complex tasks. In either case, the relationship between speed and power, whether constant (i.e., slopes constant) *or variable (i.e., slopes differ), is the crucial issue.* A reasonable prediction in the area of spatial visualization tasks would be that there are individual differences in slopes, and that the relationship between speed and level is higher for complex tasks than for simple tasks.

Varieties of Individual Differences in Spatial Ability Correlational Studies.

Given the importance of the speed-power problem, the need for a method for studying it, and the relevance of factorially designed experiments where complexity is manipulated in a systematic manner, the next step is to select the design facets. This presumes some knowledge of the important dimensions of the aptitude construct in question.

At the most basic level, spatial thinking requires (among other things) the ability to encode, remember, transform, and match spatial stimuli. Factors such as Closure Speed (i.e., speed of matching incomplete visual stimuli with their long term memory representations), Perceptual Speed (speed of matching visual stimuli), Visual Memory (short term memory for visual stimuli) and Kinesthetic (speed of making left-right, up-down discriminations) may represent individual differences in the speed or efficiency of these basic cognitive processes. However, these factors surface only when extremely similar tests are included in a test battery. Such tests and their factors consistently fall near the periphery of a scaling representation, or at the bottom of a hierarchical model.

While the processes that these factors hypothetically represent are certainly spatial in nature, they are not usually the referent of the term "spatial ability." While a number of spatial factors have been identified, only three survive a more rigorous examination using hierarchical factor methods (see Lohman, 1979a). All of the factors involve mental transformation. They are:

1. Spatial Relations. This factor is defined by tests such as Cards, Flags and Figures (Thurstone, 1938). It probably represents individual differences in the speed of mentally rotating or reflecting a simple visual stimulus. The tests that define this factor are all parallel forms of one another, and the factor emerges only if these or highly similar tests are included in the battery.
2. Spatial Orientation. This factor appears to involve the ability to imagine how a stimulus array will appear from another perspective. In the true spatial orientation test, the subject must imagine that he is reoriented in space, and then make some judgment about the situation. There is often a left-right discrimination component in these tasks, but this discrimination must be made from the imagined perspective. However, the factor is difficult to measure since tests designed to tap it are often solved by mentally rotating the array rather than reorienting an imagined self.
3. Visualization. The factor is represented by a wide variety of tests such as Paper Folding, Form Board, WAIS Block Design, Hidden Figures, Copying, etc. The tests that load on this factor, in addition to their spatial-figural content, share two important features: a) they are all administered under relatively unspeeded conditions, and b) most are much more complex than corresponding tests that load on the more peripheral factors. Tests designed to measure this factor usually fall near the center of a two dimensional scaling representation, and are often quite close to tests of Spearman's "g" (such as Raven Matrices or Figure Classification) or Cattell's (1963) Gf.

In examining the type of tests that load on these three

spatial factors, two major types of mental transformation appear to be involved. The first is mental movement. Reflecting, rotating, folding, or simply imagining that a stimulus is moved from one position in an array to another position, are all varieties of mental movement.

The second type of mental transformation may be called construction. There are two types of constructions: reproduction (i.e., physical construction) and combination (i.e., mental construction). At the simplest level, reproduction is represented in tests like Thurstone's (1938) Copying, where the subject must correctly copy a stimulus design. At the next level, it is represented by tests like Graham and Kendall's (1948) Memory for Designs, where the design must be reproduced from memory. One important difference between these tests and those that load on more peripheral factors like Perceptual Speed or Memory Span is that the design must be reproduced, not just recognized, and the reproduced design must be a veridical representation of the stimulus. Retaining a veridical mental image of a design may be an important component of other complex spatial tasks, such as Hidden Figures (French et al.,1963).

In the mental construction tasks, on the other hand, the subject must actually construct a mental image, usually by reorganizing the stimulus in a new way. The clearest examples of this sort of process are tests such as Form Equations (El Koussy, 1935) and Paper Form Board (eg., Thurstone 1938; French, Ekstrom and Price, 1963). Mental construction is an important component of many complex spatial tests. For example, in Surface Development (French et al.,1963), the task is not simply to fold the sides of the figure mentally (i.e., mental movement), but rather to construct a three dimensional figure from a two dimensional drawing. Similarly, in Paper Folding (French et al.,1963), the examinee must construct new holes as he mentally unfolds the stimulus. Finally, mental construction may take the form of mentally deleting parts of a stimulus, as in Match Problems (Guilford and Hoepfner, 1971). This may also be an important component of tests such as Embedded Figures (Witkin,

Oltman, Raskin and Karp, 1971) or Hidden Figures (French et al., 1963).

Experimental Studies of Spatial Thinking.

Until the pioneering work of Shepard and his colleagues, evidence on the nature of spatial transformation processes was limited to the speculations of factor analysts on the nature of their factors. Several important insights into the nature of spatial thinking have emerged from this work. The first significant finding was that the time required to determine whether two stimuli were identical was a linear function of the angular separation between them (Shepard and Metzler, 1971). This result was replicated in later studies using alphanumeric characters (Cooper and Shepard, 1973a, 1973b), random two dimensional polygons (Cooper and Podgorny, 1976; Cooper, 1975), and two dimensional stimuli that had to be mentally folded into three dimensional cubes (Shepard and Feng, 1972). Other important findings were: a) the function relating time to perform "in depth" rotations to the angular separation between figures was essentially the same as that for picture plane rotations (Shepard and Metzler, 1971); b) the time required to prepare for an object in a specified orientation was a linear function of the angular separation between the specified orientation and the normal or trained orientation (Cooper and Shepard, 1973a, 1973b); c) the rate of rotation for well learned complex figures was essentially the same as that for simple figures (Cooper and Podgorny, 1976); and d) when experimental conditions were equated, there were no differences between the rate of rotation for two and three dimensional stimuli (Podgorny 1975). These findings place important limitations on the type of models that may be advanced to account for the phenomenon. In particular, Shepard and his colleagues have argued that the internal processes and representations underlying the rotation task are "analog" in nature (Cooper and Shepard, 1973b; Metzler and Shepard, 1974; Shepard, 1975).

Experimental evidence on the nature of mental construction

processes is less extensive. Palmer (1974) found that the time to synthesize a stimulus mentally was inversely related to the "goodness" of the component parts. However, there were important individual differences in discriminative reaction time to the test stimulus, indicating that some of the subjects may not have completely synthesized the component parts. Glushko and Cooper (1976) used a different paradigm to investigate mental construction. They asked subjects to construct a mental image of a composite figure on the basis of a verbal description of the component parts. Construction time increased with the number of component parts, but discriminative reaction time to a test figure was independent of figural complexity.

Individual Differences in the Experimental Studies.

Although most experimental paradigms relegate individual differences to the error term, individual differences in performance on spatial tasks have been recognized to be striking enough that investigators are now forced to deal with them. Even within small, select samples, the estimated rates of mental rotation for individual subjects have differed by a factor of two or three (Cooper and Shepard, 1976). There are important individual differences in discriminative reaction time to the test stimulus as well. Cooper (1976) (see also Cooper and Podgorny, 1976) found that, for some subjects, "same" responses were, on the average, faster than "different" responses. Further, the speed of "different" responses was not related to the similarity of the test stimulus to the standard. For other subjects, "different" responses were generally faster than "same" responses, and "different" reaction time decreased as the standard and test stimuli became increasingly dissimilar. Cooper (1976) speculates that these patterns may reflect important differences in the nature of the comparison processes.

Finally, there are important individual differences in the ability to solve the more difficult spatial tasks in the first place. For example, of the ten Stanford undergraduates who participated in the Shepard and Feng (1972) experiment, three had

error rates so high (16-29%) that their data were discarded. Further, the correlation (for the seven remaining subjects) between the number of errors and mean reaction time on the 40 error free items was -.28, which is consistent with the previous argument here regarding the relative independence of speed and level.

Solution Strategies.

Individual differences in solution strategy are a major problem for both experimental and correlational studies of spatial thinking. While some subjects solve verbal problems such as anagrams, syllogisms, and three term series problems using a predominately spatial strategy, it is possible to construct verbal tasks where spatial strategies would be of little or no assistance (e.g., a simple vocabulary test). On the other hand, it is extremely difficult to devise spatial tasks that cannot be solved by some nonspatial strategy.

Further, subjects often change strategies as they become more familiar with the task, or as item difficulty increases (Meyers, 1958; Barrat, 1953; Lohman, 1977a). Careful selection of the experimental task and open response format can eliminate some strategy shifts. Although retrospective and introspective reports are often of dubious validity, they can be obtained, and may suggest at least major individual differences in strategy or radical shifts in strategy over items.

However, relationships between individual differences in patterns of scores over facets and reference constructs can provide more direct evidence on the nature of overall strategies and shifts in strategy over items. For example, the correlation between the pattern of latency cell means over a rotation facet and reference variables may suggest that steep slopes are related to verbal ability and, by implication, verbal strategies. Other mean contrasts may be formulated to address both general and specific hypotheses about strategies, such as:

1. Do changes in difficulty effect overall changes in speed-accuracy tradeoff?

2. What is the relationship between individual differences in speed-accuracy tradeoff, overall or over a particular facet, and reference variables?
3. What are the relationships between individual differences in the patterns of cell means for latency or correctness over a particular facet and reference variables?

Examination of the relationships between individual differences in the patterns of scores over design facets and reference tests can provide a powerful method for exploring intra- and inter-subject variation in solution strategy (Calfee, 1976).

Methodological Issues

Limitations of Tests and Factors.

Since correlations between task parameters and reference tests and factors constitute a major portion of the results of this experiment, some comments on the advantages and limitations of such data are in order. While it is important to demonstrate that performance on the experimental task is related to performance on reference tests or factors, such correlation (or lack of it) cannot be considered the final arbiter of construct validity. There are good tests and bad tests. Many tests, and especially instruments intended only for research (e.g., French, Ekstrom and Price, 1963) are not carefully constructed, even by psychometric standards. Further, most, if not all, test items can be solved in more than one way. This is particularly true of spatial tests. For example, only seven subjects in this experiment reported that they always rotated the stimuli on the Cards and Figures tests. Five reported that they never rotated, while most (13) reported rotating some stimuli and using logical cues to solve other items. Data were unavailable for six subjects. Thus, it is unreasonable to expect large correlations between performance on these tests and performance on rotation items in the experimental task.

Similarly, factors representing common covariation in

clusters of tests can be shadow or substance. Some of the more specific factors (e.g., Spatial Relations, Closure Speed, Perceptual Speed) are the product of including what amounts to a parallel forms reliability coefficient in the matrix of test intercorrelations. Such factors regularly disappear when the tests are altered to remove method covariance. Further, factor analysis assumes that subjects are not solving tests in different ways. Correlations and factor patterns can change drastically when data are analyzed within distinguishable strategy groups (French, 1965; Lohman, 1979a).

Thus, correlations between total scores, cell totals, or individual contrast scores and reference tests can provide useful but limited evidence on the construct validity of the score or contrast. The importance of reference tests lies in the link, no matter how weak or uncertain it may be, with the network of individual difference constructs established through fifty years of correlational research on human abilities.

Two more specific limitations of the reference test correlations should also be noted. First, the 30 subjects who participated in this experiment were deliberately selected to represent a wide range of verbal and spatial abilities. Correlations between task parameters and reference tests could be quite different in a more restricted sample.

Second, subjects apparently interpreted the instruction to "solve each item as quickly as you can" in quite different ways. While the type of analysis performed provided considerable insight into the effects of speed-accuracy tradeoffs, the results of the experiment could change substantially if subjects were somehow induced to perform as rapidly as they could without sacrificing accuracy.

Finally, it must be emphasized that this experiment was designed to investigate the speed-level hypothesis. Thus, even though considerable evidence on the nature of individual differences in spatial thinking emerged from the analysis, the study was not designed to investigate such hypotheses rigorously.

Contrast Correlations

When subject-factor interactions are large, plots of mean scores may be misleading. This is especially true when the subject-factor interaction is large and significant, and the factor itself is nonsignificant.

Subject variability in the dependent measure over a particular design facet may be examined by computing contrast scores for each subject and then correlating these scores with overall performance and reference tests (Calfee, 1976). Examples of procedure for computing individual contrast scores are presented elsewhere (see Lohman, 1979b).

Conclusions.

Previous research on the nature of aptitude processes has ignored the important differences between speed and level of performance. In particular, the speed of correctly solving simple items is largely independent of level attained on more complex items of the same type. Failure to recognize this fact has produced considerable confusion in the factor analytic literature, and in factor models of human abilities. Likewise, it has resulted in constructs, models, and parameters in experimental psychology that are of questionable generalizability, particularly for process understandings of constructs such as "verbal ability," "spatial ability," and "intelligence."

The key to the resolution of the speed-power problem is the construction of psychologically meaningful indices of item difficulty. This is perhaps best accomplished by equating difficulty with complexity of processing requirements, and then constructing information processing models that account for the various dimensions of task complexity.

Within the area of spatial abilities, the two major complexity facets are construction and mental movement. Construction may be divided into physical construction (i.e., reproduction) and mental construction. Mental movement includes rotation, reflection, folding, and transposition. Complex spatial tasks usually require more than one of these processes,

such as folding and mental construction. Other complex spatial tasks require either the repeated application of a single process (e.g., mental rotation), or the execution of one process on a complex, unfamiliar stimulus.

Overview of the Study.

The present experiment was designed to explore individual differences in spatial ability, with particular emphasis on the interrelationships of speed, level and complexity, and their relationships with aptitude reference constructs. Complexity was represented in two ways: stimulus complexity and processing complexity. Stimulus complexity was determined by the number of points on both the given and to-be-constructed stimuli. Processing complexity was manipulated by increasing the "amount" of a particular operation (i.e., degrees of rotation or number of stimulus pieces that are added together) as well as the variety of operations (construction and rotation alone versus construction plus rotation).

Mental movement was represented by mental rotation, in the tradition of Shepard and his colleagues (Shepard and Metzler, 1971; Cooper and Podgorny, 1976). Construction was represented in two ways: a) by requiring the subject to combine two or three stimulus figures mentally, and b) by asking him to draw the combined, rotated, or combined and rotated stimulus figure.

The major hypotheses of the investigation were:

1. Speed of performing simple spatial tasks is largely independent of the level of such tasks that a subject can ultimately solve.
2. Level scores correlate more strongly than corresponding speed scores with higher-order aptitude constructs such as G, Gf, and Vz.

The present study also provided an opportunity to study the relationships between individual differences in the patterns of latency and correctness cell means over design facets and reference aptitude constructs.

CHAPTER II
DESIGN AND PROCEDURE

Design

There were four major components in the design of this experiment: the experimental task, the draw trials, the simple reaction time control, and the reference constructs.

Experimental Task

The design of the experimental task is shown in Figure 2.1. All design facets were fully crossed except level one of the construction facet, which was not crossed with type of addition or product complexity.

Levels of the construction facet refer to the number of stimulus figures presented during a trial. When more than one stimulus was presented, later stimuli were to be added to the initial stimulus in a prescribed manner to construct a new stimulus.

Levels of the rotation facet indicate the number of degrees through which the subject was required to rotate the given or constructed stimulus. When both construction and rotation were required on a given trial, construction always preceded rotation.

Stimulus complexity refers to the average number of points on the stimulus figures presented during a trial. The levels were low (3 - 4 points), medium (5 - 6 points), and high (7 - 8 points).

Match refers to whether a correct or incorrect test stimulus was presented at the conclusion of the trial.

Type of addition refers to whether the to-be-added stimuli were added to the right (level 1) or the left (level 2) of the initial stimulus.

Product complexity refers to the number of points on the constructed stimulus. Again, low meant 3 or 4 points and high 7 or 8 points.

Since the entire design was replicated within each subject, subjects may be considered the seventh fully crossed factor.

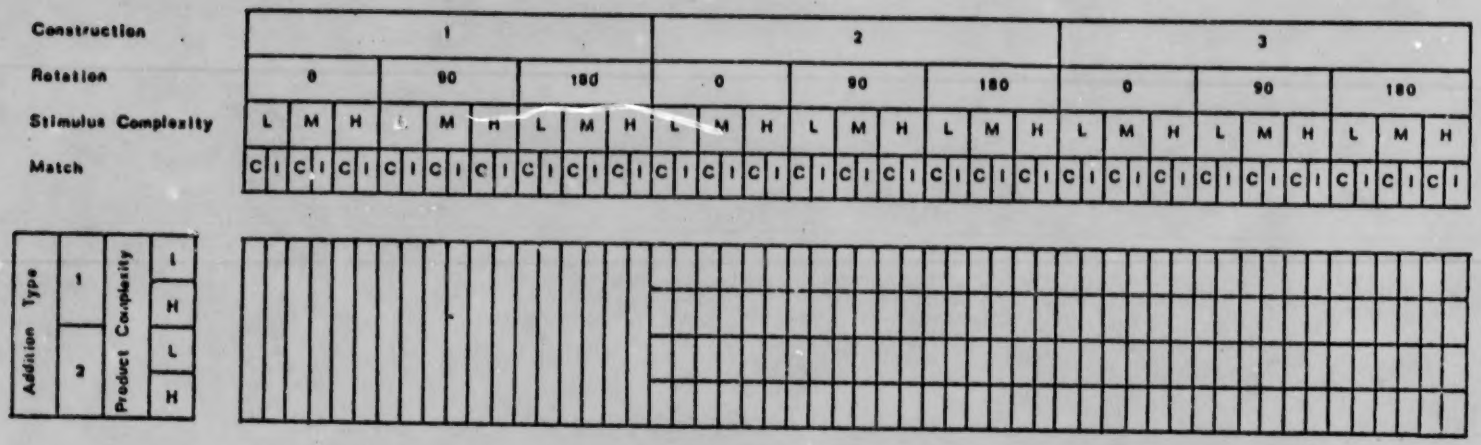

Figure 2.1 Design of the experimental task.

Draw Trials

The draw trials were chosen by taking a one quarter fraction of a 2^5 factorial design in which the match facet was eliminated and all three level factors were converted to two level factors. This was accomplished by ignoring level one of the construction facet, defining a rotate and no rotate factor and ignoring level 2 (medium complexity) on the stimulus complexity factor.

The quarter fraction of this design defined 16 cells. One item was constructed for each no rotation cell and two for each rotation cell, one for both 90 and 180 degree rotations. Further, magnitude of rotation was counterbalanced with the four values of product complexity (3,4,7,8) within each level of the construction facet.

In addition to these 24 items (eight from the no rotation cells and 16 from the rotation cells), six items were constructed to represent the crossing of rotation and stimulus complexity at level one of the construction facet. The complete set of draw trials thus contained 30 items.

Simple Reaction Time Control

This condition consisted of three consecutive blocks of 20 true-false color discrimination trials. The ten true and ten false trials were separately randomized within each block.

Reference Constructs

Since the major goal of the study was to examine individual differences in spatial thinking, the selection reference constructs constituted an important aspect of the overall design. Tests representing general intelligence, Gf, Gc, Visualization, Perceptual Speed, Closure Speed and Memory Span were already available for these subjects. Additional tests were chosen to tap particular aspects of individual differences in spatial thinking not clearly represented in any of the previous reference tests. Cards (French, Ekstrom and Price, 1963) and Figures (Thurstone, 1938) were chosen to represent the Spatial Relations factor. Copying (French, Ekstrom and Price, 1963) and Memory for

Designs (Graham and Kendall, 1948) were chosen to represent rudimentary aspects of spatial construction (see Chapter I). A complete list of the reference battery, with a brief description and example item from each test is reported in Lohman (1979b).

Assignment of Items to Blocks

The experimental task was organized into five blocks. The first block contained the 60 reaction time control trials. The next four blocks were composed of experimental and draw trials. The 216 experimental trials were first randomly assigned to one of the four blocks under the constraint that all main effects be kept independent of block. Since there was only one observation per cell at levels 2 and 3 of the construction facet, some confounding of interaction vectors and block was inevitable. The degree of confounding was determined by correlating vectors representing the linear and quadratic components of blocks (in the order administered to all subjects) with the design vectors. For both linear and quadratic block vectors only five correlations with design vectors were greater than .20. For the linear component, only one correlation was above .25 (KRST2, r=.27). For the quadratic component, two correlations were above .25 (RSMP1, r=.27; and KRSMTP2, r=.27). Thus, even for the design vectors most strongly correlated with block, only 7% of the variance accounted for by those vectors may be attributed to linear or quadratic effects over blocks.

The 30 draw trials were also randomly assigned to one of the four blocks, for totals of 62, 61, 61, and 62 items in blocks one through four respectively. Items were then randomly ordered within each block before filming.

Subjects

Subjects were selected from an initial reference population of 241 Palo Alto high school students and 123 Stanford University undergraduates. A large reference battery was administered to these subjects in the Spring of 1976. Most had participated in other experiments conducted by the Aptitude Project. Of the

original 171 males, 158 were available for the present experiment. Those who had participated in similar experiments were eliminated. Thirty-one subjects were then selected from the remaining sample on the basis of composite spatial and verbal scores. Tests were selected for each of the three composites on the basis of a within sex factor analysis of the entire reference battery. The first spatial composite was formed by normalizing and then summing scores for Paper Folding, Surface Development, Hidden Figures (all from French, Ekstrom and Price, 1963), Block Design, Picture Arrangement (both from Wechsler, 1955) and the advanced Raven Progressive Matrices (Raven, 1962). This composite is akin to Cattell's (1963) Gf or Zimmerman's (1953) Vz. The second spatial composite was formed by normalizing and then summing each subject's scores on Paper Form Board (French, Ekstrom and Price, 1963) and Object Assembly (Wechsler, 1955). Although the factor analytic evidence for the separation of these two tests from those in the first composite was weak, both tests were hypothesized to involve a form of mental construction that could be important in the present experiment. A scatterplot of scores on the two spatial composites for the available population of males is shown in Figure 2.2. Subjects selected for the present experiment are represented by squares in this figure. The correlation between the two composites was .68.

The verbal composite was formed by summing normalized scores for an abbreviated version of the Terman Concept Mastery Test (Terman, 1950; however, see Lohman, 1979b), Information, Comprehension, Similarities and Vocabulary (all from Wechsler, 1955).

A scatterplot of scores on the first spatial composite and the verbal composite for the 158 males is shown in Figure 2.3. Again, subjects participating in the present experiment are represented by squares.

As can be seen in Figures 2.2 and 2.3, the subjects who participated in this experiment were quite representative of the reference population on both spatial ability and verbal ability.

All subjects were paid for their participation. One subject

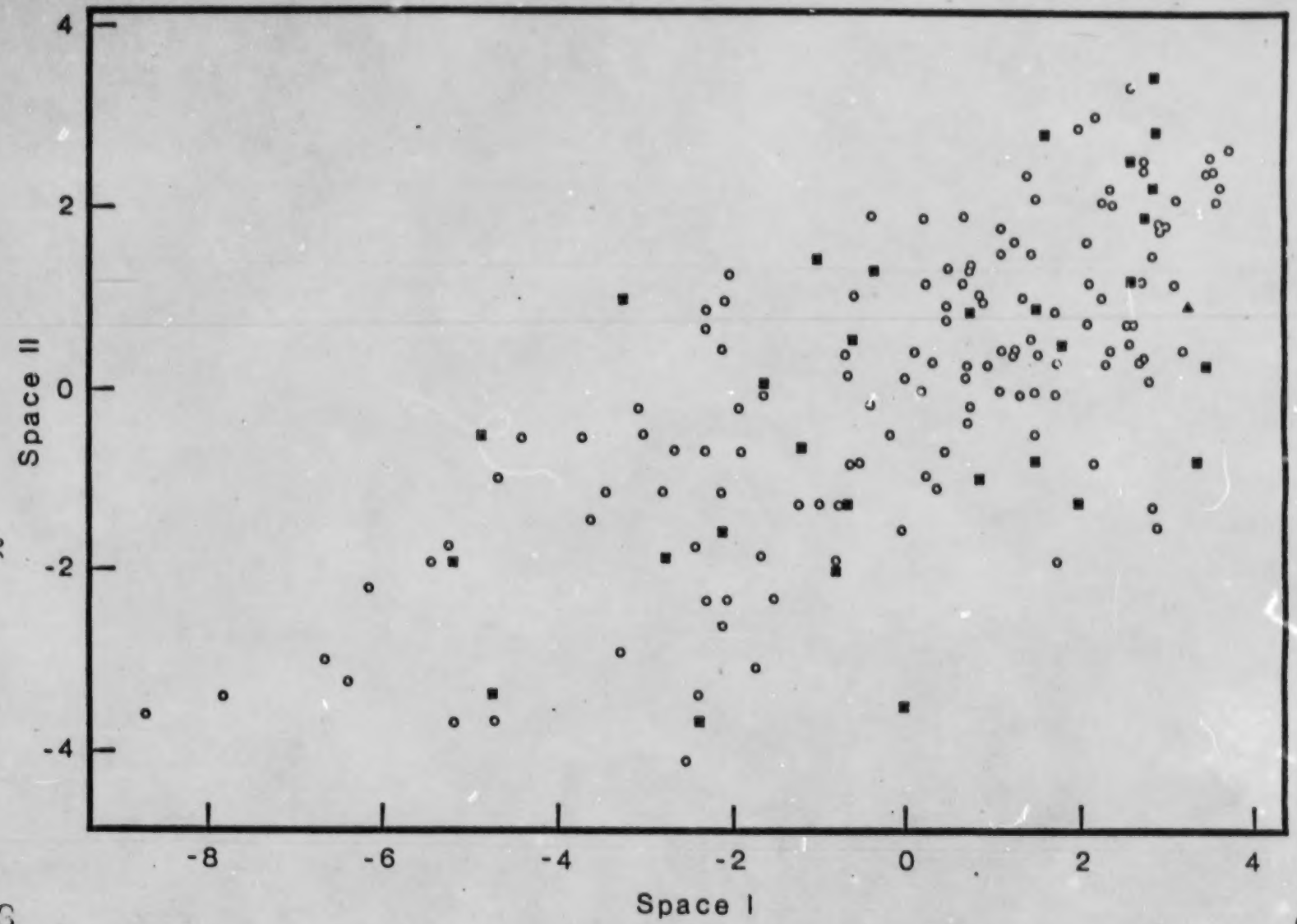

Figure 2.2. Scatterplot of scores on the two spatial composites for all 158 males in the reference population.

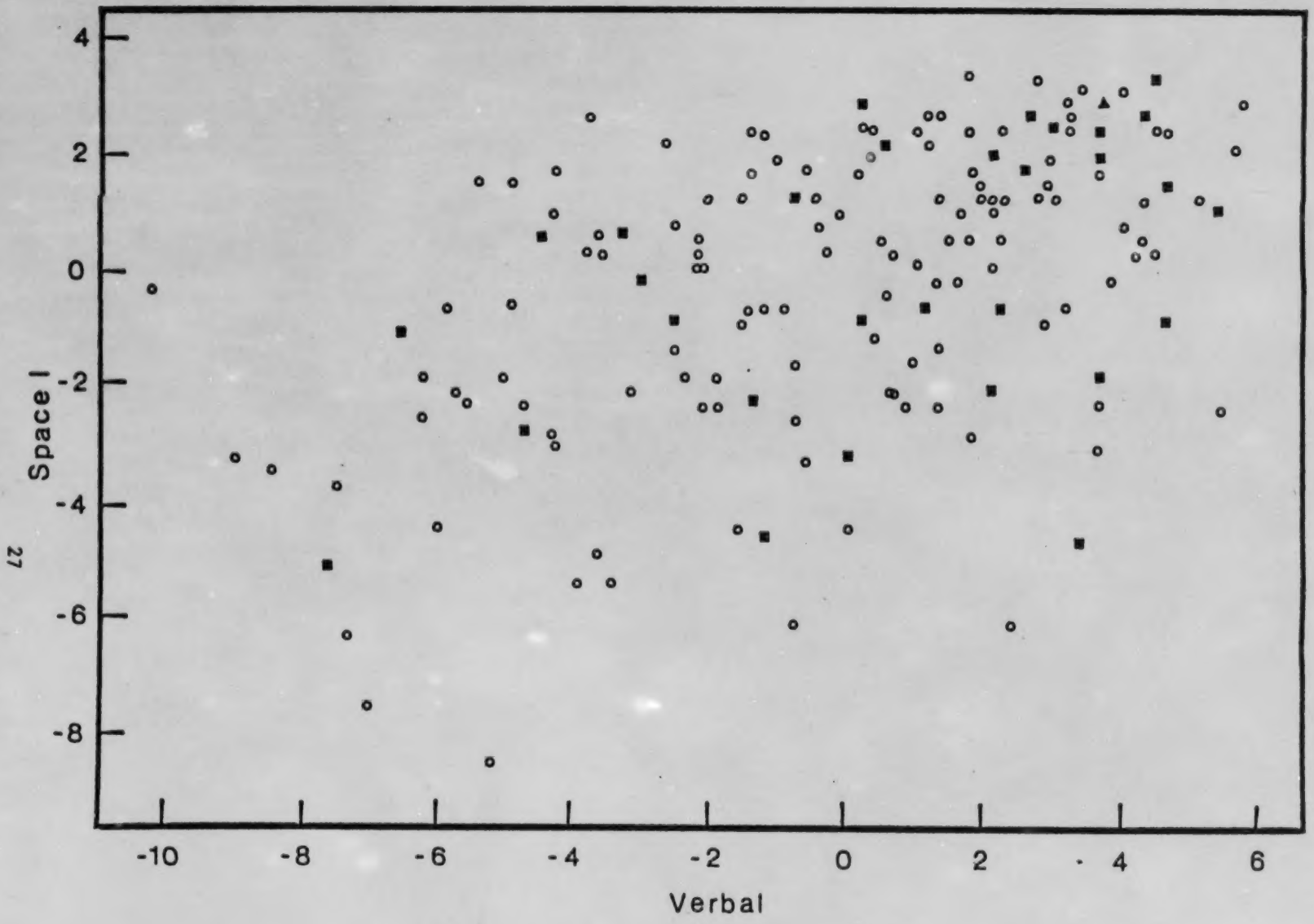

Figure 2.3. Scatterplot of scores on the first spatial and verbal composites for all 158 males in the reference population.

completed only the first session of the experiment and had to be replaced. His scores are represented by a triangle in Figures 2.2 and 2.3.

Stimuli

Overview of Trial Types

Trials differed in the number, variety and magnitude of transformations required, the type of test probe, and the complexity of the given or constructed stimulus. Ignoring for the moment type of test probe, stimulus complexity and product complexity, the major variants of trial types are shown in Figure 2.4. For purposes of illustration, all test stimuli are correct.

As is shown in Figure 2.4, from four to six frames were presented on each trial. The first and last frames were always "ready" and "confidence" respectively. The first stimulus was presented on the second frame, and the test stimulus or "draw" frame immediately preceded the "confidence" frame.

Column one in Figure 2.4 shows the sequence of events when no rotation or construction were required: first the "ready" frame, next the presentation of the base stimulus then the test probe, and finally the confidence frame.

Columns two and three in Figure 2.4 show the sequence of frames for 90 and 180 degree rotations. Type one addition is shown in columns four and five, and type two addition in columns six and seven. Note that when two stimuli are to be added on the second frame, the stimulus on the left must be added first. Also note that the "+" sign indicates the location of the previous stimulus. Two examples of the four possible combinations of both construction and rotation are shown in columns eight and nine. Finally, column ten presents an example of a draw trial. Draw trials were the same as any other trial except that an instruction to draw the resulting figure replaced the test probe.

Construction of Stimuli

Seven sets of 75 stimuli were obtained by randomly selecting and then plotting three to eight pairs of points in each of seven

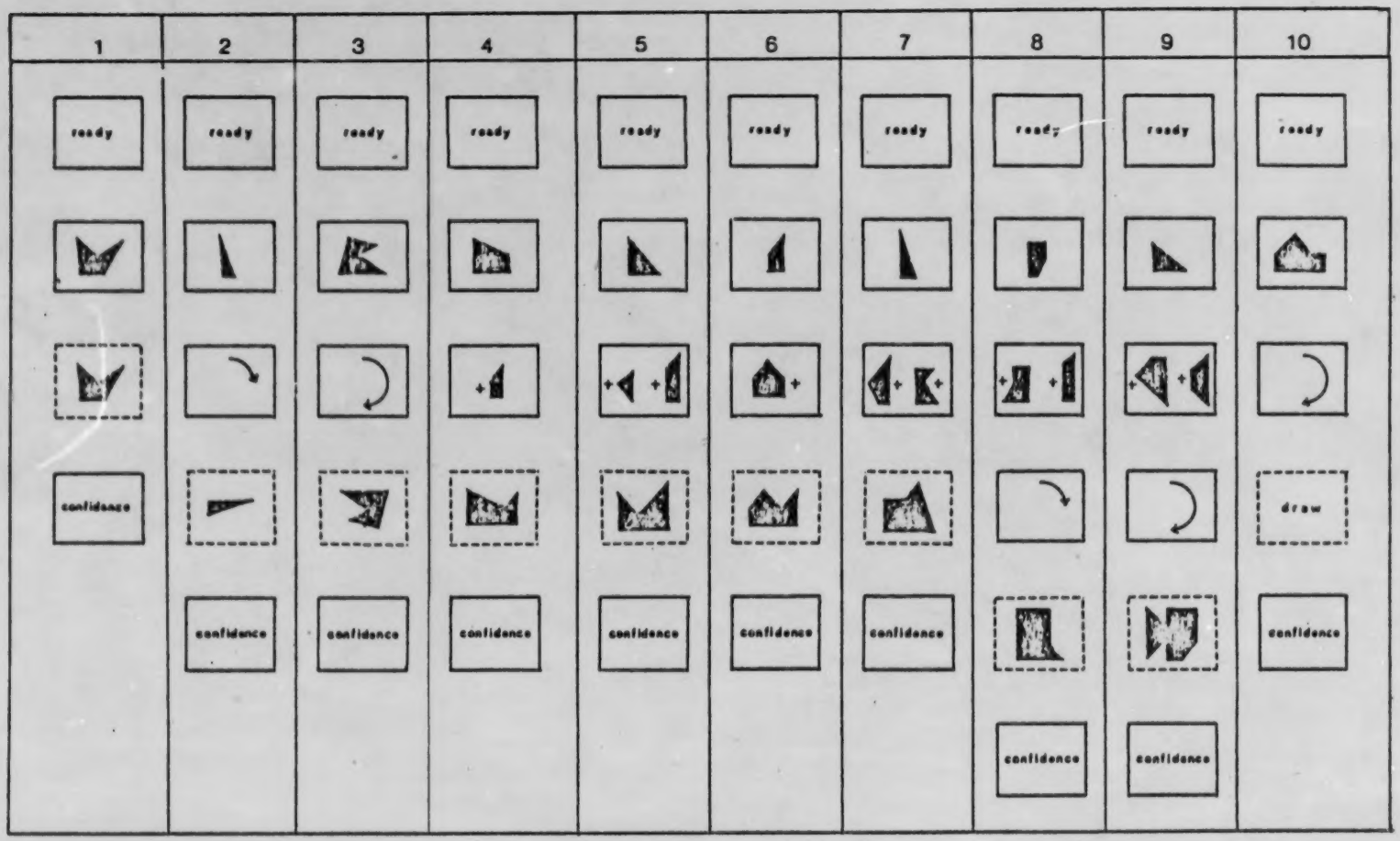

Figure 2.4 Examples of the major trial types in the experimental task.

different 3x3 to 8x8 grids. The points were then connected using the Method 1 of Attneave and Arnoult (1956). An informal pilot study with these stimuli revealed that most were far too difficult to remember or manipulate, especially when attempting to combine them mentally. In particular, irregular forms were much more difficult than regular or symmetric forms. Accordingly, irregular stimuli were "normalized" by making almost parallel lines, parallel, nearly symmetric lines, symmetric, angles and lines that were nearly equal, equal, and angles that were nearly 90 degrees, right angles. Four to eight point stimuli without at least one right angle, or two parallel or symmetric lines were then eliminated. Each figure was then oriented so that either one line or the major axis was vertical or horizontal. Examples of stimuli from each of the three levels of complexity are shown in Figure 2.5. Since all three point stimuli are triangles, stimulus complexity was necessarily confounded with stimulus familiarity (or uniqueness) in the design. Every effort was made to avoid using the same stimulus more than once. Thus, even though all three point stimuli were triangles, there are an infinite number of different triangles. Some repetition of forms was unavoidable, particularly on those items where low complexity stimuli were combined to form a low complexity product. Incorrect test stimuli were formed by reflecting or incorrectly rotating either the entire stimulus or one of its component stimuli.

Construction of Construction Items

Construction items were developed by choosing a target stimulus of the required complexity (low or high), and then cutting it into the appropriate number of pieces of average complexity equal to that required by the cell (low, medium or high). Low, medium, or high complexity stimuli could be combined to form either a low or high complexity product. Note that the level of stimulus complexity was defined as the average number of points on the two or three to-be-combined stimuli.

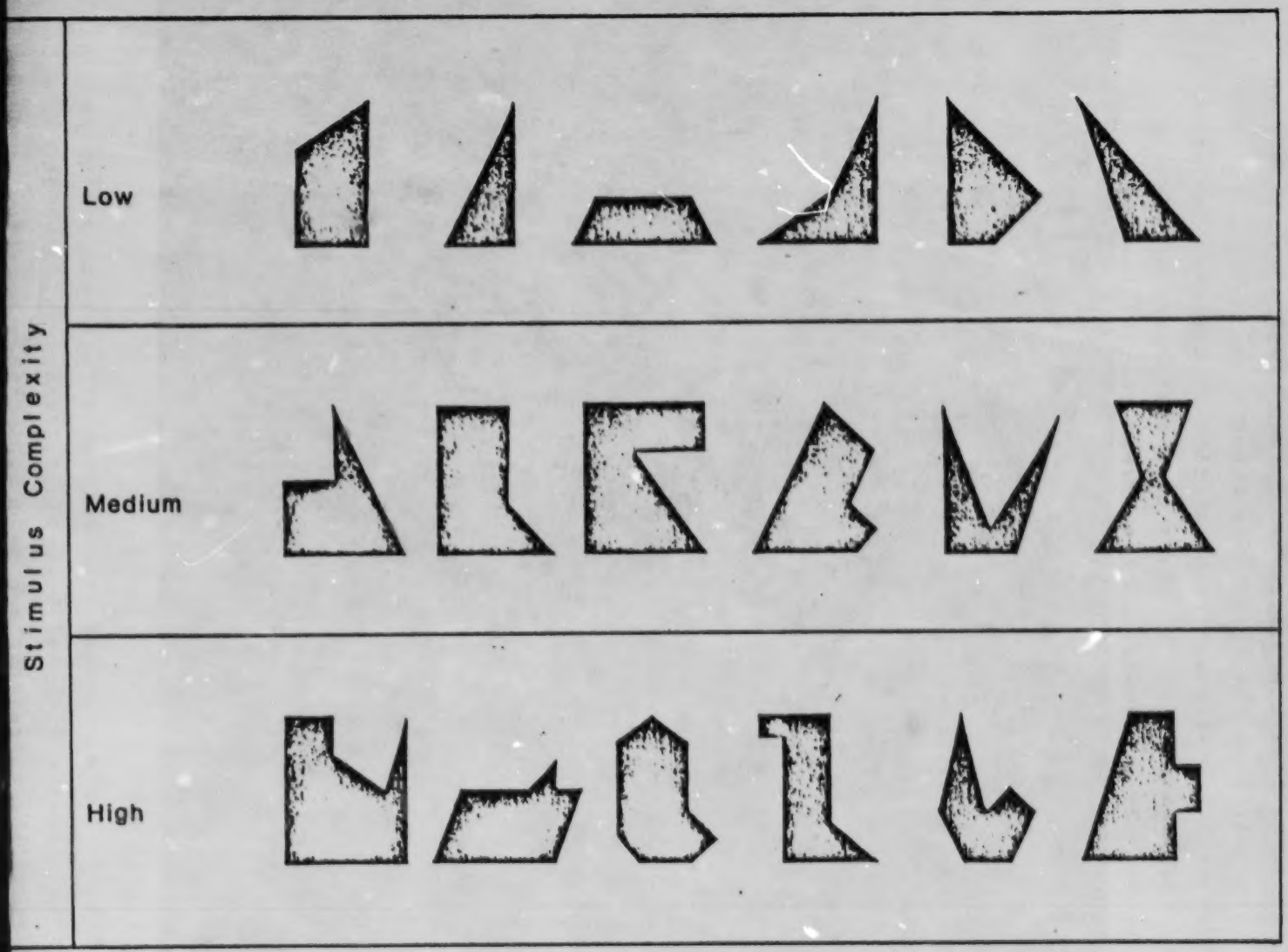

Figure 2.5. Examples of stimuli at each of the three levels of stimulus complexity.

Apparatus

All stimuli were photographed on Kodak ASA 40 Super 8 film for presentation on a single frame Super 8 projector. Stimuli appeared as black silhouettes on a white background. Rotation arrows and construction "+" signs were also black on a white background. Test stimuli and the "draw" instruction were silhouetted on a yellow background. The background on the ready and confidence frames was red. The film was projected on a 21 x 28 cm rear screen. Average stimulus figure size within this field was approximately 8 cm. The distance between subject and screen was approximately 60 cm. Subjects controlled the presentation of stimuli by pressing either the hand held "true" and "false" buttons, or one of the four confidence buttons mounted on a small keyboard in front of the subject.

Pressing any one of these buttons would cause the projector to advance to the next frame. A South West Technical Products 6800 microprocessor recorded which key had been pressed, the elapsed time on the frame to the nearest hundreth of a second, and then advanced the projector to the next frame. The change from one frame to the next was virtually instantaneous.

Procedure

Each subject participated in two sessions, each lasting approximately two hours. In the first session, subjects were given a general orientation to the experiment. The four reference tests were then administered in the following order: Cards (four minutes/part), Copying (three minutes/part), Figures (eight minutes total), and Memory for Designs. The 15 stimuli in the Memory for Designs test were drawn on three by five inch cards and presented individually for five seconds. Subjects then drew each design after it was removed. They were told that the designs would be scored both for accuracy and correct proportions.

The 60 simple reaction time control trials were given next. Each trial began with a red background "ready" slide which remained on the screen for 4.6 seconds. Either a yellow or white

slide then appeared. Subjects were told to answer the question "Is it yellow?" as rapidly as possible by pressing one of the hand held buttons. Subjects pressed the "yes" with their preferred hand. Each subject was first given 10 practice trials and then the 60 test trials.

After a ten minute break 11 demonstration experimental trials were given. Subjects were told to work as rapidly as possible but to be sure they had a clear image of the stimulus before advancing to the next frame. They were told that a yellow background test probe could occur at any point during the item. Further, they were told that if they took longer than one second to respond to the test probe they would not receive credit for the item.

The demonstration was followed by nine practice items. On both practice and demonstration, questions, important distinctions and errors were explained by having the subject physically combine and rotate duplicate cardboard stimuli.

Finally the first block of experimental trials was presented. Subjects were asked to rate their confidence in each answer on a four point scale (zero = guessing; three = certain). They were asked to tell the experimenter if they accidentally pressed "yes" when they meant "no" or vise versa, and to press "no" if they could not perform a given construction or rotation. Finally, they were told that their drawings would be scored for both accuracy and correct proportions.

In the second session, subjects were first given ten practice trials from the first block, and then blocks two, three and four with a 10 minute break between blocks. At the conclusion of the experiment subjects were asked to explain how they were solving the items. They were then thanked and paid three dollars per hour for their participation.

CHAPTER III
RESULTS AND DISCUSSION

Overview

This chapter reports the major results of the experiment. It is divided into six sections. The first section reports preliminary analyses on the formation of the various dependent measures, removal or replacement of outliers, computation of individual guessing scores, and the formation of a new index of correctness.

The second section reports a series of regression analyses on the major speed-level hypothesis. These analyses ignore the design facets and examine the relationships between speed and level using a psychometric index of item difficulty.

The third section reports the separate means analyses and correlational analyses on total time, correctness, and confidence. Here, the effects of the design facets on latency and correctness are examined in greater detail. The effect of practice on latency and correctness is first examined briefly. Then, the problem of reaction times for incorrect items is discussed and analyses on correct and all reaction times are reviewed. Separate analyses on total time and correctness are then reported in some detail. The interpretation of correlations between individual contrast scores and reference tests is discussed, and the contrast correlations for the separate analyses presented. The results of the total time and correctness analyses are then compared.

In the fourth section, joint analyses on total time and correctness are presented and compared with the separate analyses. It is argued that the joint indices provide unique insights for a means analysis, but may distort the correlational analyses.

In the fifth section, separate analyses on the encoding, construction, rotation, and match component times are reported. The relationships between component times and scores, and specific reference factors such as Closure Speed and Perceptual

Speed are also discussed.

In the sixth and final section, the major results of the total time and correctness analyses are compared with the results of the component time and correctness analyses.

I. Preliminary Analyses

Dependent Measures

Three types of scores were obtained on each trial: time, correctness and confidence. Time on each frame was recorded to the nearest hundredth of a second. Total trial time was determined by summing the component times, excluding, of course, the "ready" and "confidence" frame times.

Two correctness measures were obtained. The first represented the correctness of the subject's response to the test (match) stimulus. The second measure was at the component level. Subjects were requested to press the "no" button if they were unable to perform a particular component operation.

At the conclusion of the item, subjects rated their confidence in their response to the test stimulus on a zero (guessing) to three (certain) point scale.

Outliers

Outliers were identified by examining scatterplots of total and component times versus composite design vector indices. This permitted an examination of the distribution of scores within each cell of the design. For example, the abcissa in the total time scatterplots represented a linear ordering of cells according to construction, rotation, stimulus complexity, and product complexity.

These scatterplots revealed that an arbitrary statistical rule for defining outliers would be unworkable. Times that were extremely deviant for the simplest item types were not necessarily extreme within the total distribution. Only .3 to .7 percent of the scores were identified as outliers by this method. These scores were either eliminated, or replaced by the next highest time in the cell, depending on the analysis.

Computation of Individual Guessing Scores

On any rating scale, some subjects utilize the entire scale, while others concentrate on one end of the scale. Thus, a new variable called GUESS was calculated for each subject by computing the probability of a correct response at each level of reported confidence. Guess was then defined as the level of confidence where the probability of a correct response was not greater than .5 at an alpha of .10.

A More Stringent Definition of Correctness

Component times become uninterpretable if subjects do not perform the operations indicated on a given frame before advancing to the next frame. The random presentation of items, the unpredictability of the test probe, the draw trials, and the instruction to respond to the test probe within one second were all designed to minimize this problem.

A number of different factors, including failure to perform component operations when requested, may result in an unusually long discrimination reaction time. The distribution of these "match" times was produced for each subject. All showed some degree of bimodality. The lowest point between the two peaks was then used to define a new variable called TMATCHN (N for Node) for each subject. These node times ranged from 1.6 to 3.2 seconds, with a median of 2.0 seconds. An example of the distribution of match times and the location of TMATCHN for one subject is shown in Figure 3.1.

These new indices were used to define an additional criterion of item correctness called SCORE3. An item was counted correct only if it was correct (i.e. SCORE = 1), if confidence was greater than GUESS, if there were no "no" responses to component operations, and if match time was less than TMATCHN.

II. Regression Analyses

The relationships between latency and correctness may be examined in two ways. The first ignores the design facets, is more atheoretic, briefer and resembles the traditional

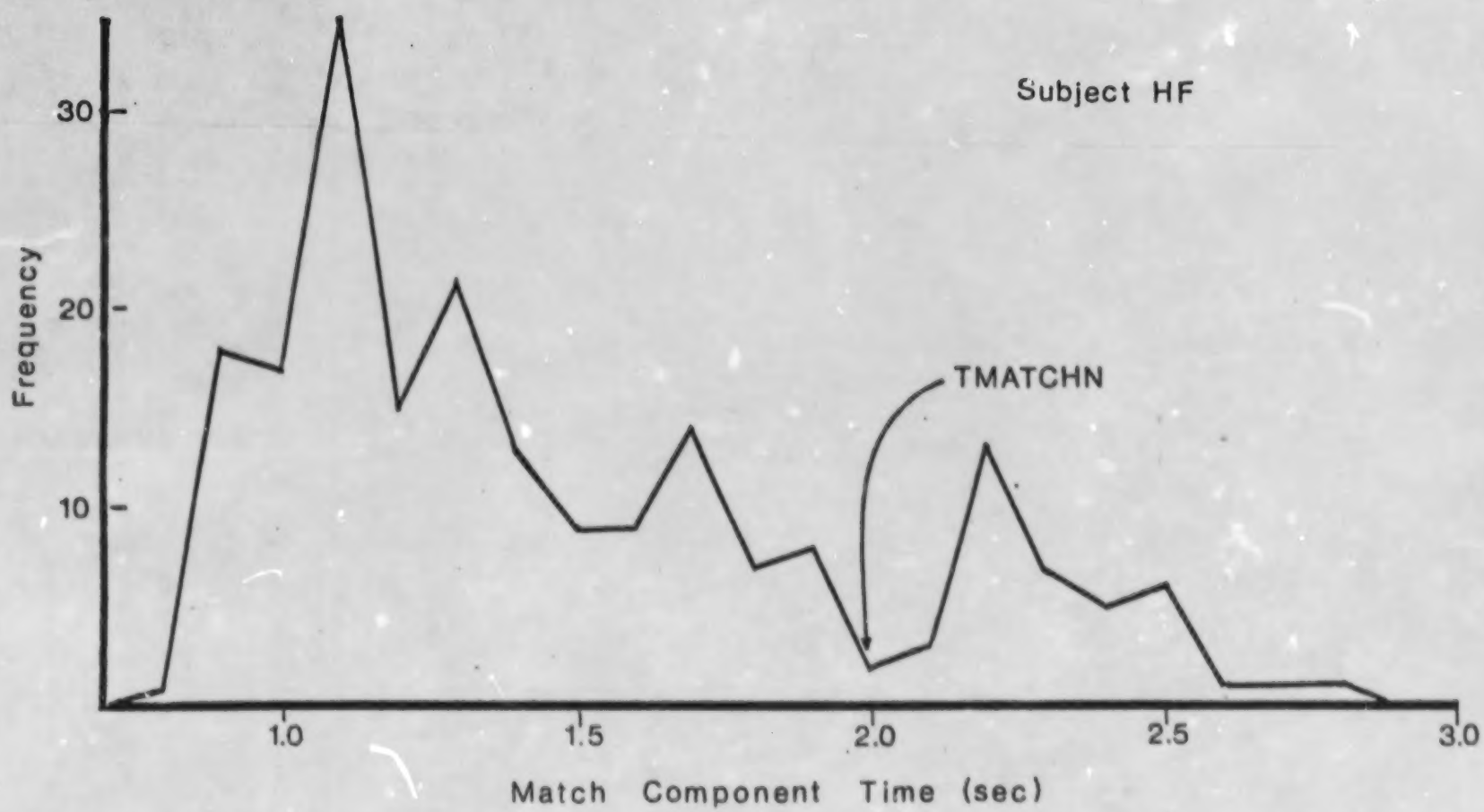

Figure 3.1. Distribution of match component times and the location of TMATCHN for subject HF.

psychometric type of analysis. The second examines the relationships between latency and correctness in the light of the design facets, assumes that these facets affect differential changes in performance, is more complex, but also more enlightening. This section reports four analyses of the first type. Analyses of the second type constitute the remainder of the chapter.

Construct Validity of the Level Scores

Before examining the speed-level hypothesis, it is first necessary to establish the construct validity of the level score. Correlations between total SCORE, total SCORE3, total correct Draw trials, and the major reference tests are shown in Table 3.1. Total SCORE correlated .75 with Paper Folding and .66 with Memory for Designs-Quality. The Draw total correlated .81 with Paper Folding, .77 with Surface Development, and .73 with Memory for Designs-Quality. Further, Draw and total correct intercorrelated .82. Clearly, individual differences in level on this task were congruent with individual differences in spatial ability on the reference tests.

All three variables had their highest correlations with spatial tests. With two notable exceptions, the spatial tests correlations for SCORE and SCORE3 were virtually identical. SCORE correlated higher with Paper Folding, while SCORE3 had the higher correlation with Figures. The crucial difference appeared in the correlations with verbal tests: SCORE had uniformly higher correlations with verbal tests than SCORE3.

To clarify these relationships, two clusters of reference tests were formed. The spatial cluster was composed of Paper Folding, Form Board, Surface Development, Cards, Figures and Memory for Designs-Quality. The verbal cluster contained WAIS Information, Comprehension, Similarities, Vocabulary, and Word Beginnings and Endings. SCORE and SCORE3 were then correlated with the sum and the difference of these two clusters. The verbal plus spatial composite represents general mental ability (in fact, it correlated .99 with WAIS Full Scale Score), while

Table 3.1

Correlations between Total SCORE, Total SCOTE3, Total Correct Draw Trials, and Reference Tests (N=30)

Reference Test	Total SCORE	Total SCORE3	Total Draw
Paper Folding	75	57	81
Form Board	48	45	58
Surface Development	66	61	77
Copying	42	32	57
Cards	49	41	50
Figures	45	54	46
Memory for Designs	39	41	53
Memory for Designs - Quality	66	63	73
Visual Number Span	40	42	43
Identical Pictures	39	26	50
Word Transformations	36	25	43
Camoflauged Words	40	04	48
Hidden Figures	53	30	52
Word Beginnings and Endings	41	21	52
Necessary Arithmetic Operations	47	21	52
Letter Series	67	41	55
Terman Concept Mastery	51	27	64
Raven Progressive Matrices	45	18	49
WAIS Information	48	26	54
Comprehension	33	25	55
Similarities	46	38	56
Digit Span	51	44	36
Vocabulary	50	24	61
Digit Symbol	27	15	47
Block Design	48	23	50
Picture Arrangement	35	41	42
Object Assembly	28	29	51
Verbal Scale Score	53	34	64
Performance Scale Score	52	39	69
Full Scale Score	60	41	76

Note. Decimals omitted.

the spatial minus verbal composite represents the spatial-verbal distinction. If SCORE3 is a "cleaner" spatial score, it should correlate higher with the spatial end of the spatial-verbal composite than SCORE. The correlations with the sum composite were .73 and .64, and with difference composite .29 and .42 for SCORE and SCORE3, respectively. Thus, SCORE3 is a cleaner measure of spatial ability than SCORE.

Speed-Level Correlations

Items were first rank ordered according to percent of the sample failing the item by SCORE3. Total item time was then regressed on percent failing, separately for each subject. The analysis was then repeated using only those times for items correct by SCORE3. Since items were ordered on a zero to one scale, the intercept provided the best estimate of speed of solving the simplest item types. Predicted speed of solving the most difficult items was estimated by simply summing the slope and intercept. Correlations between intercepts, slopes, predicted times at maximum difficulty for the two analyses, and reference spatial tests are shown in Table 3.2. Table 3.3 contains a similar set of correlations, this time excluding the scores for one subject who had an extremely steep slope.

The most important correlations are those between the three variables that represent level on this task (SCORE, SCORE3, Draw) and the intercepts and predicted times at maximum difficulty. In all four analyses, the intercepts were independent of the level scores. On the other hand, predicted times at maximum difficulty were highly correlated with total SCORE and total Draw, but unrelated to total SCORE3.

Examination of the correlations with reference tests explain this discrepancy between the SCORE and SCORE3 correlations with predicted times for difficult items. In all four analyses, the intercept correlated highest with a strategy index, number of pencil marks each subject made on the Form Board test. The correlations were strongest in the analyses on correct times. When all times were included, the correlation with Closure Speed

Table 3.2

Correlations between Slopes, Intercepts, Predicted Times at Maximum Difficulty, and Reference Tests (N=30)

Reference Test	All Times			Correct Times Only		
	Intercept	Slope	$\hat{T}$	Intercept	Slope	$\hat{T}$
Paper Folding		40	35			
Form Board	-33					
Form Board - Marks	41			54		
Surface Development	-38	32				
Copying	-36					
Cards						
Cards - Wrong		49	45		59	53
Figures						
Closure Speed	-45					
Visual Number Span		34	39		33	38
Identical Pictures						
Object Assembly	-40	35		-31		
Memory for Designs- Quality						
Word Beginnings & Endings		47	40		36	32
Necessary Arith. Operations		44	39		40	37
Terman Concept Mastery		54	53		50	49
Information		46	44		44	43
Comprehension		44	36		42	35
Similarities	-36	38		-30		
Digit Span		46	39		32	
Vocabulary		54	49		47	44
Verbal Scale Score		58	50		50	44
Performance Scale Score	-33					
Full Scale Score	-32	52	42		40	34
Draw Total		58	53		45	44
Total SCORE		56	56		42	46
Total SCORE3						

Note. Decimals and correlations less than .30 omitted.

Table 3.3

Correlations between Slopes, Intercepts, Predicted Times at Maximum Difficulty, and Reference Tests (N=29)

Reference Test	All Times			Correct Times Only		
	Intercept	Slope	$\hat{T}$	Intercept	Slope	$\hat{T}$
Paper Folding		53	46		42	37
Form Board	-34	36				
Form Board - Marks	41		37	53		30
Surface Development	-38	48	32		33	
Copying	-38					
Cards						
Figures						
Closure Speed	-47			-34		
Visual Number Span		39	44		44	47
Identical Pictures		40	32			
Hidden Figures		50	50		40	41
Word Beginnings & Endings		48	38		35	
Necessary Arith. Operations		44	36		42	35
Letter Series		43	47		41	45
Terman Concept Mastery		52	48		49	46
WAIS Information		43	41		46	42
Comprehension		30				
Similarities	-36	37				
Digit Span		46	37			
Object Assembly	-40					
Verbal Score		55	43		46	36
Performance Score	-33	37				
Memory for Designs-Quality		61	56		59	54
Draw Total Correct		70	61		62	55
Total SCORE		69	66		59	58
Total SCORE3		40	33		32	

Note. Decimals and correlations less than .30 omitted.

became significant. When the outlier was removed (Table 3.3), correlations between the intercept and other spatial tests surfaced, although none reached significance.

The slope, on the other hand, had its highest correlations with verbal tests in the full sample analyses (Table 3.2). Here, when all times were included, slope correlated highest with WAIS Verbal Scale Score, Terman Concept Mastery, WAIS Vocabulary, and number of errors on the Cards test. Later results will show that Cards Wrong is actually an index of verbal strategy, not spatial ability. When only correct times were included in the analysis, slope correlated exclusively with verbal tests and Cards Wrong.

When the outlier was removed, the correlations presented more of a mixed bag. Here (see Table 3.3), slope correlated highest with Memory for Designs - Quality, WAIS Verbal Scale Score, Terman Concept Mastery, and Hidden Figures. However, on balance, the correlations were still stronger with verbal tests than spatial tests.

In sum, the major hypothesis of this investigation is supported by these results, but with an unexpected twist. Speed of solving simple items was unrelated to level on the task. Speed of solving difficult items was highly correlated with two of the level scores. However, it was expected that these predicted times at maximum difficulty would be negatively correlated with spatial reference tests. Instead, predicted times for difficult items were positively correlated with verbal reference tests.

Subsequent analyses on correctness and latency will clarify these relationships between verbal ability, spatial ability, and speed and level scores on the experimental task. This is accomplished by examining the structure of cell means, through analysis of variance, and the relationships between individual differences in patterns of scores over design facets and reference tests, through contrast and cell total correlations.

III. Separate Analyses on Total Time, Correctness and Confidence. Practice Effects

For the group data, there was a significant decrease in total time over the four blocks. Mean times averaged over subjects and items within a block were 19.46, 16.18, 14.92, and 13.06 sec for blocks one through four, respectively. An analysis of variance indicated that while the linear term captured most of the variance ($F = 255.51$) the quadratic and cubic terms were also significant ($F = 6.09$, $p = .0138$ and $F = 4.16$, $p = .042$, respectively).

At the individual level, nine subjects showed no change or minimal change in RT over blocks, thirteen showed a definite decline in RT, and eight had other patterns (such as a large drop between block one and two, but steady thereafter, or a gradual increase in RT over blocks).

Subjects also tended to make fewer errors with practice. The percentage of correct responses was .78, .80, .83 and .83 for blocks one through four respectively. The corresponding values for SCORE3 were .60, .67, .72, and .72.

RT for Incorrect Items

When a subject fails to solve an item, the reaction time for that item is ambiguous. While some investigators regularly include both correct and incorrect reaction times in their analyses (e.g., Sternberg, 1977), the usual procedure is to exclude times from incorrect items (Pachella, 1974). However, this experiment differs from the classical RT study in two important ways:

1. Subjects were deliberately selected to represent a wide range of individual differences in spatial ability. Using only reaction times for correct items would bias the results by weighing more heavily the performance of the more able subjects.

2. Similarly, items were constructed in order to represent a wide range of item difficulty or complexity. An analysis of correct reaction times

would be biased, giving more weight to the easier items.

In any case the issue of whether to include only the "clean" correct RT's or all RT's in an analysis is an empirical matter. As a preliminary check, an analysis of variance was performed on all reaction times (except 30 outliers, which were excluded) and on the subset of these times where the item was correct and confidence was greater than GUESS. The results are reported elsewhere (see Lohman, 1979b). Note that these are not repeated measures analyses of variance. Computer limitations precluded doing this type of analysis on the incomplete data matrix defined by the second analysis. Therefore, all subject factor interactions were included in the error term. Both analyses followed a regression model in which main effects were entered first, two way interactions second, etc.

Including all times in the analysis added some noise. The squared multiple correlations were .352 and .419 for the complete and restricted data sets, respectively. A few interactions were significant in one analysis and nonsignificant in the other, but on the whole, the results were quite similar. This does not mean that there are not important psychological differences between correct and incorrect reaction times. Rather, it merely indicates that including all times in the analysis of variance did not seriously alter the results.

Analysis on Total Time

Means Analysis

A repeated measures analysis of variance on total time is shown in Table 3.4. Times for both correct and incorrect items were included in the analysis. Thirty extreme observations were replaced by the next highest within cell time.

All main effects were in the expected direction, except match, which was not significant. Construction, rotation, and stimulus complexity all had strong linear components and weak, but significant quadratic components. These main effects are plotted in column one of Figure 3.5, below. Note also that there

Table 3.4

Analysis of Variance on Total Time

Source	df	MS	F	Source	df	MS	F
Mean	1	1624939.49	263.10*	N	29	6176.11	-
Construction	(2)			NC	(58)		
Linear	1	131612.26	127.81*	Linear	29	1029.76	33.72*
Quadratic	1	2089.17	64.61*	Quadratic	29	32.34	1.06
Rotation	(2)			NR	(58)		
Linear	1	33597.19	121.55*	Linear	29	276.40	9.05*
Quadratic	1	2295.83	35.29*	Quadratic	29	65.05	2.13*
Stimulus Complexity	(2)			NS	(58)		
Linear	1	22628.14	54.38*	Linear	29	416.12	13.63*
Quadratic	1	298.99	6.01	Quadratic	29	49.78	1.63
Match	1	23.66	.69	NM	29	32.28	1.12
Type of Addition	1	2787.89	47.45*	NT	29	58.75	1.92*
Product Complexity	1	16867.77	38.58*	NP	29	437.18	14.32*
CR	4	359.27	11.11*	NCR	116	32.34	1.06
CS	4	1009.25	22.55*	NCS	116	44.75	1.47*
CM	2	15.62	.79	NCM	58	19.76	.65
KT	1	3003.60	38.51*	NKT	29	78.00	2.55*
KP	1	4026.37	31.07*	NKP	29	129.58	4.24*
RS	4	118.33	4.01*	NRS	116	29.49	.97
RM	2	276.42	5.88*	NRM	58	47.02	1.54*
RT	2	976.06	16.76*	NRT	58	58.24	1.91*
RP	2	1006.13	22.32*	NRP	58	45.08	1.48
SM	2	46.96	1.48	NSM	58	31.75	1.04
ST	2	1175.67	18.06*	NST	58	65.09	2.13*
SP	2	1067.99	10.33*	NSP	58	103.38	3.39*
MT	1	632.59	13.23*	NMT	29	47.83	1.57
MP	1	1195.24	11.80*	NMP	29	101.26	3.32*
TP	1	2834.35	26.24*	NTP	29	108.02	3.54*
CRS	8	124.37	2.87	NCRS	232	43.29	1.42*
CRM	4	378.73	8.69*	NCRM	116	43.58	1.43*
CSM	4	566.70	11.91*	NCSM	116	47.59	1.56*

Table 3.4 (continued)

Source	df	MS	F	Source	df	MS	F
KRT	2	806.78	23.73*	NKRT	58	34.00	1.11
KRP	2	189.15	5.39*	NKRP	58	35.07	1.15
KST	2	1093.37	16.80*	NKST	58	65.09	2.13*
KSP	2	1215.51	32.46*	NKSP	58	37.45	1.23
KMT	1	298.62	2.80	NKMT	29	106.69	3.49*
KMP	1	28.17	.55	NKMP	29	51.59	1.69
KTP	1	1446.99	15.82*	NKTP	29	91.48	3.00*
RSM	4	714.50	18.92*	NRSM	116	37.77	1.24
RST	4	408.47	8.37*	NRST	116	48.80	1.60*
RSP	4	347.32	5.01*	NRSP	116	69.37	2.27*
RMT	2	1007.53	31.00*	NRMT	58	32.50	1.06
RMP	2	2547.56	19.97*	NRMP	58	127.56	4.18*
RTP	2	261.53	10.66*	NRTP	58	24.54	.80
SMT	2	152.07	2.37	NSMT	58	64.07	2.10*
SMP	2	773.87	11.92*	NSMP	58	64.92	2.13*
STP	2	101.94	3.05	NSTP	58	33.37	1.09
MTP	1	71.89	1.38	NMTP	29	52.09	1.71
CRSM	8	212.70	6.05*	NCRSM	232	35.17	1.15
KRST	4	1325.56	16.46*	NKRST	116	80.54	2.64*
KRSP	4	1649.01	15.25*	NKRSP	116	108.11	3.54*
KRMT	2	142.35	2.91	NKRMT	58	49.00	1.60*
KRMP	2	1370.03	15.85*	NKRMP	58	86.42	2.83*
KRTP	2	158.55	3.94	NKRTP	58	40.20	1.32
KSMT	2	281.82	3.85	NKSMT	58	73.28	2.40*
KSMP	2	110.36	1.85	NKSMP	58	59.55	1.95*
KSTP	2	300.06	5.50*	NKSTP	58	54.55	1.79*
KMTP	1	45.83	.90	NKMTP	29	51.08	1.67
RSMT	4	110.71	2.51	NRSMT	116	44.04	1.44*
RSMP	4	514.25	7.94*	NRSMP	116	64.80	2.12*
RSTP	4	117.99	3.32	NRSTP	116	35.54	1.16
RMTP	2	765.32	10.81	NRMTP	58	70.77	2.32*
SMTP	2	933.65	13.07*	NRMTP	58	71.45	2.34*

Table 3.4 (continued)

Source	df	MS	F	Source	df	MS	F
KRSMT	4	53.22	.90	NKRSMT	116	59.22	1.94*
KRSMP	4	396.04	5.78*	NKRSMP	116	68.48	2.24*
KRSTP	4	433.76	8.55*	NKRSTP	116	50.71	1.66*
KRMTP	2	447.89	7.33*	NKRMTP	58	61.07	2.00*
KSMTP	2	482.70	3.63	NKSMTP	58	132.86	4.35*
RSMTP	4	796.33	9.71*	NRSMTP	116	82.02	2.69*
KRSMTP	4	1237.93	11.33*	NKRSMTP	116	109.22	3.58*
				Residual	1620	30.54	-

*p less than .01

were many large interactions. Thus, a simple additive model does not fit the data well.

All two-way interactions except CM and SM were significant. The eight largest of these are plotted in Figure 3.2. The construction by rotation interaction indicates that the time taken to rotate and match a constructed stimulus was slightly greater at 180 degrees for two piece additions than one piece additions.

Although the type of addition main effect was significant, the construction by type (KT) interaction reveals that left additions took longer only for two piece additions. Similarly, the type of addition by product comlexity (TP) interaction reveals that left and right additions at low product complexity took the same amount of time, while left additions were much more time consuming than right additions at high product complexity.

The construction by product complexity (KP) interaction reflects the fact that high product complexity additions were relatively more time consuming for two piece additions than one piece additions. Note that in all interactions between construction and type of addition or product complexity, construction is represented by K. This is necessary in order to distinguish between the two level construction contrast between two piece additions and one piece additions (K), and the C linear and quadratic contrasts that span the entire three levels of the construction facet. Discussion of the RT and RP interactions is best postponed until after the contrast correlations are presented and correctness and latency are examined jointly.

Objections to Contrast Correlations

Two apparent limitations of the correlations between these contrast scores and reference variables merit brief comment. First, correlations based on only 30 cases are usually quite unstable. However, subjects were not selected randomly; rather they were deliberately chosen to be a representative sample of verbal and spatial ability in the larger population (N = 171 high school and college males). Therefore, the correlations reported

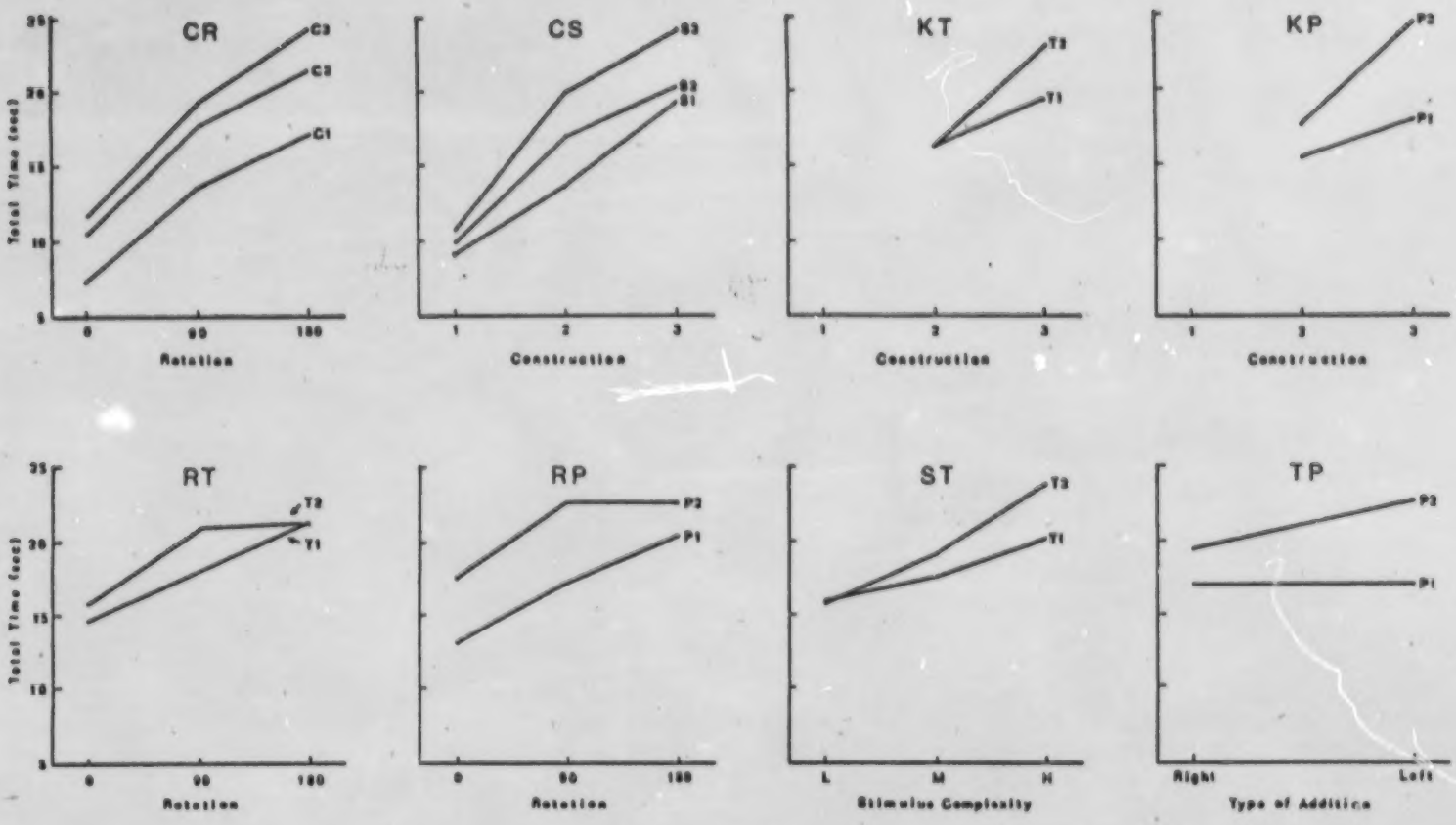

Figure 3.2 Plots of the eight largest two way interactions for total time.

here are much more dependable than the sample size suggests.

Second, all contrasts except the N contrast are difference scores. Difference scores are known to be relatively unreliable (Lord, 1963; Cronbach and Furby, 1970). Five factors mitigate these complaints.

1. Each contrast score is based on at least 144 observations and as many as 216 observations for each subject. Thus, the means and differences between them should be quite stable.
2. As Cronbach and Furby (1970) point out, linked errors can substantially enhance the reliability of difference scores. Since all observations were obtained from one task there necessarily are linked errors in these scores. Hence, the reliability of the difference scores is greater than would be estimated by conventional methods (i.e., Lord, 1963).
3. All reliability coefficients are merely estimates of score reliability. Low estimated reliability may help explain why a score fails to correlate with other variables. However, if the reliability estimates are low, and the correlations are high, then it is the reliability estimate that needs explaining. Further, even within the rigid assumptions of classical test theory, it is the square root of the estimated reliability that defines the limit in validity. Thus, if the reliability of a test is only .25, the test may (in theory) correlate .50 with another variable.
4. It can be shown (see Lohman, 1979b) that the correlation between the difference score X-Y and another variable Z will be greater than the difference between r(X,Z) and r(Y,Z) if r(X,Y) is greater than .50. If r(X,Y) is .80, then r(X-Y,Z) = 1.58 [r(X,Z)- r(Y,Z)]. If r(X,Y)=.95, then r(X-Y,Z) = 3.16 [r(X,Z)-r(X,Y)]. Thus, when cell

intercorrelations are high, as in this task, correlations between individual contrast scores and reference tests may be quite high. Further, since the contrast score correlation with a reference test will be much larger than the difference between corresponding correlations between each cell and the reference test, it is much easier to detect differential correlations between performance over a task and reference tests using the contrast correlations.

5. In most educational applications, the absolute magnitude of the gain score is frequently about the size of the standard error of measurement of the raw scores, and much smaller than the range of individual differences in the raw scores. However, the magnitude of the difference score over a powerful facet in a within subjects design may be much larger than the standard error of the raw scores and even larger than the the range of individual differences in the raw scores.

Thus, contrast scores may be more dependable and more useful than is commonly assumed in the psychometric literature on difference scores. Nevertheless, there are three important limitations of contrast scores that must be noted:

1. Even though contrast scores may be more dependable than the psychometric literature suggests, they are still less reliable than their corresponding raw scores.
2. Ceiling and floor effects in either raw score can destroy the validity of the difference score.
3. While analysis of variance procedures are relatively robust with respect to violations of the equal variance assumptions, correlations are not. Thus, correlations between contrast scores and other variables can be seriously biased if the within cell variances are not approximately equal. In particular, such correlations will tend to reflect the rank order of raw scores in the cell with the greater variance.

Total Time Contrast Correlations

Significant correlations between reference tests and the ten total time main effect contrasts are summarized in Table 3.5.

N Contrast. The N contrast represents subject means. Mean total time correlated .49 with total SCORE but only .18 with total SCORE3. Therefore, those who take longer tend to get more items correct. However, when scores are corrected for guessing and extremely long match times, mean time and correctness become independent. The only other significant N contrast correlation was .43 with Terman Concept Mastery. Correlations with other verbal tests were positive, but not significant (i.e., .37 with WAIS Information, and .35 with Vocabulary), while correlations with spatial tests were all much smaller, and, of course, not significant. Thus, if anything, mean time appears to be more related to verbal ability than spatial ability.

Construction Contrasts. The construction linear contrast represents the difference between average time on two piece additions and no addition items. Scores on this contrast were most consistently correlated with verbal tests and number of errors on the Cards test. The quadratic contrast represents the difference between the average of no additions plus two piece additions, and one piece additions. High positive scores reflect little change between zero addition and one addition, and a large increase between one and two additions. Negative scores reflect the reverse pattern. The contrast correlated negatively with Visual Number Span (r=-.45) and Hidden Figures (r=-.51) and positively with Marks Visual Imagery Questionaire (r=.43). Therefore, those who reported vivid imagery were more likely to show the most change between no additions and one addition, while those with high scores on Hidden Figures and Visual Number Span showed larger increases between one and two piece additions. The most important finding here was the opposition between Hidden Figures, which is probably the most "analytic" spatial test in the battery, and the reported vividness of visual imagery. This suggests that an analytic approach to the construction task can

Table 3.5

Correlations between Total Time Main Effect Contrast Scores and Reference Tests (N=30)

N		C Linear		C Quadratic		R Linear		R Quadratic	
Total SCORE	49	Cards Wrong	52	Marks Imagery Q.	43	Cards Wrong	57	Picture Compl.	54
Terman Concept M.	43	Total Draw	46	Vis. Numb. Span	-45				
		Vis. Numb. Span	45	Hidden Figures	-51				
		Terman Concept M.	53						
		Information	44						
		Vocabulary	47						
		WAIS Verbal S.	47						
S Linear		**S Quadratic**		**M**		**T**		**P**	
WAIS IQ	50	WAIS IQ	64	Form Board	-44	WAIS IQ	52	WAIS IQ	60
Paper Folding	44	Paper Folding	47	Cards Wrong	65	Paper Folding	48	Paper Folding	57
Terman Concept M.	44	Identical Picture	53	Vis Numb Span LC	51	Aud Lett Span LC	43	Form Board	46
Comprehension	48	Finding A's	50	Picture Compl.	-49	Word Transform.	48	Surface Develop.	50
Digit Span	51	Word Begin/End	54			Terman Concept M.	52	Mem for Designs-Q	57
Vocabulary	51	Nec. Arith. Oper.	47			Information	50	Identical Picture	45
WAIS Verbal S.	60	Terman Concept M.	47			Vocabulary	46	Information	43
		Information	42			Object Assembly	62	Comprehension	42
		Comprehension	51			WAIS Verbal S.	47	Similarities	43
		Similarities	57			WAIS Perform. S.	46	Digit Span	49
		Digit Span - F	61					Vocabulary	56
		WAIS Verbal S.	61					Object Assembly	49
		WAIS Perform. S.	52					WAIS Verbal S.	59
								WAIS Perform. S.	48

Note. Decimals and correlations less than .42 omitted.

be distinguished from a more holistic imagery approach. The two variables were associated with different patterns within the task, even though neither was significantly related to overall performance.

Rotation Contrasts. There were only two significant correlations between the rotation contrasts and reference tests. The linear contrast correlated .57 with errors on the Cards test, and the quadratic contrast correlated .54 with WAIS Picture Completion. Later results suggest that errors on the Cards test are related to verbal strategies. Thus, those who showed the greatest time differences between no rotation and 180 degree rotation may have been using nonspatial strategies to rotate the stimuli.

Stimulus Complexity. Both the linear and quadratic stimulus complexity contrasts correlated strongly with verbal ability. The linear contrast correlated significantly with seven reference tests, and only one of these was a spatial test (Paper Folding, r=.44). The highest correlation was with WAIS Verbal Scale Score (r=.60).

The stimulus complexity quadratic contrast correlated significantly with fouteen reference tests. Ten of these were verbal or verbal reasoning tests, three were perceptual speed tests, and one was a spatial test (Paper Folding, r=.47). The highest correlations were .61 with WAIS Verbal Scale Score and Digit Span Forward. Thus, those who showed the largest increase in total time between low and medium complexity stimuli had low scores on these reference tests, while those who showed the greatest increase between medium and high complexity stimuli scored highest on these reference tests.

Match. Although the match facet was not significant in the analysis of variance, it did have four large correlations with reference tests. Scores on this contrast represent the difference between time taken to accept a correct test stimulus

and the time required to reject an incorrect stimulus. Thus, positive scores reflect longer acceptance times and negative scores longer rejection times. The contrast correlated negatively with Form Board and Picture Completion, and positively with Cards Wrong and Visual Number Span. Therefore, those who were faster accepting a correct probe scored higher on Form Board and Picture Completion, while those who were faster rejecting an incorrect probe made more errors on the Cards test and scored higher on the Visual Number Span test. Both of these latter variables were associated with verbal strategies throughout the analysis. Form Board and Picture Completion, on the other hand, are good examples of spatial tests that place a premium on comparing a constructed or remembered stimulus with a given stimulus. Therefore, as suggested in the comparison of SCORE and SCORE3, match times may provide a useful index of solution strategy.

Type of Addition. Individual scores on this contrast correlated significantly with eight reference tests. The highest correlation was .62 with WAIS Object Assembly. Thus, those who obtained higher scores on this test took much longer on left additions than on right additions. Six correlations (ranging from r=.43 to .52) were with verbal tests, and one with another spatial test (Paper Folding, r=.48). Therefore, excepting the strong correlation with Object Assembly, time once again was related to verbal ability.

Product Complexity. Contrast scores on this facet correlated with both spatial tests and verbal tests. Therefore, the effect is most clearly associated with general ability. High ability subjects took much longer on high complexity products than low complexity products. Further, this contrast was the best predictor of overall performance on this task. The correlation with total SCORE was .65, and with Total Correct Draw trials .73.

Analyses on Correctness

Means Analysis.

Analyses of variance for SCORE and SCORE3 are shown in Tables 3.6 and 3.7. Estimated variance components and percent of variance associated with each facet in the SCORE3 analysis are reported elsewhere (see Lohman, 1979b). The data were first collapsed over the match and type of addition facets, except at level one of construction, where match was treated as replications. Each cell thus contained four observations. Cell means were then computed with two means per cell at level one of the construction facet and one mean at levels two and three. Replications within the level one construction cells provided an estimate MS(e) used to test the subject-facet interactions.

The two analyses produced quite different results. In particular, the linear contrast for stimulus complexity was highly significant in the SCORE3 analysis, but nonsignificant in the SCORE analysis. Product complexity was also highly significant in the SCORE3 analysis and only marginally significant in the SCORE analysis. On the other hand, the stimulus complexity by product complexity interaction was much stronger in the SCORE analysis.

The major difference between these two variables is that a subject could obtain credit for an item under SCORE when he guessed, had little faith in his answer or took an inordinately long time to accept or reject the test stimulus. SCORE3 attempted to remedy these deficiencies. Correlations between total SCORE, total SCORE3, and reference tests previously examined (see Table 3.1) indicated that SCORE3 was a cleaner measure of spatial ability. The analyses of variance on the two variables confirm this. SCORE3 was a better measure of spatial ability than SCORE on two counts. Its internal validity, as measured by congruence with the model, and external validity, as measured by differential correlations with reference tests, were both higher.

Plots for the construction, rotation, stimulus complexity, and product complexity SCORE3 main effects are shown in the

Table 3.6

SCORE Analysis of Variance

Source	df	MS	F	Source	df	MS	F
Mean	1	1051.747	2921.52*	N	29	.360	-
Construction	(2)			NC	(58)		
Linear	1	4.896	102.01*	Linear	29	.048	2.78*
Quadratic	1	2.044	40.87*	Quadratic	29	.050	2.90*
Rotation	(2)			NR	(58)		
Linear	1	.994	16.56*	Linear	29	.060	3.47*
Quadratic	1	.081	2.32	Quadratic	29	.035	2.03*
Stimulus Complexity	(2)			NS	(58)		
Linear	1	.058	3.63	Linear	29	.016	.93
Quadratic	1	.005	.14	Quadratic	29	.036	2.08*
Product Complexity	1	.363	6.85	NP	29	.053	3.07*
CR	4	.166	5.01*	NCR	116	.033	1.93*
CS	4	.240	6.73*	NCS	116	.034	2.07*
KP	1	.063	1.30	NKP	29	.049	2.84*
RS	4	.048	1.73	NRS	116	.028	1.59*
RP	2	.023	.72	NRP	58	.032	1.85*
SP	2	.359	9.32*	NSP	58	.038	2.23*
CRS	8	.086	2.86*	NCRS	232	.030	1.73*
KRP	2	.029	.76	NKRP	58	.038	2.20*
KSP	2	.620	14.95*	NKSP	58	.041	2.40*
RSP	4	.254	7.63*	NRSP	116	.033	1.93*
KRSP	4	.204	5.78*	NKRSP	116	.035	2.04*
				Residual	270	.017	-

* = p less than .01

Table 3.7

SCORE3 Analysis of Variance

Source	df	MS	F	Source	df	MS	F
Mean	1	740.133	1030.82*	N	29	.718	-
Construction	(2)			NG	(58)		
Linear	1	14.118	160.43*	Linear	29	.088	3.34*
Quadratic	1	4.129	74.44*	Quadratic	29	.057	2.17*
Rotation	(2)			NR	(58)		
Linear	1	2.827	27.19*	Linear	29	.104	3.95*
Quadratic	1	.039	.88	Quadratic	29	.044	1.67
Stimulus Complexity	(2)			NS	(58)		
Linear	1	1.387	33.82*	Linear	29	.041	1.56
Quadratic	1	.000	.01	Quadratic	29	.024	.91
Product Complexity	1	5.703	64.80*	NP	29	.088	3.34*
CR	4	.270	6.35*	NCR	116	.042	1.62*
CS	4	.284	7.33*	NCS	116	.039	1.47*
KP	1	.029	.49	NKP	29	.059	2.24*
RS	4	.046	1.23	NRS	116	.038	1.43
RP	2	.198	3.97	NRP	58	.050	1.90*
SP	2	.206	4.13	NSP	58	.050	1.90*
CRS	8	.135	3.49	NCRS	232	.039	1.47*
KRP	2	.100	2.83	NKRP	58	.036	1.35
KSP	2	.789	15.00*	NKSP	58	.053	2.00*
RSP	4	.209	4.51*	NRSP	116	.046	1.76*
KRSP	4	.603	14.79*	NKRSP	116	.041	1.55*
				Residual	270	.026	-

* = p less than .01

second column of Figure 3,.5 (below). Both rotation and stimulus complexity show linear declines in correctness. The important contrast for construction was between no addition (i.e., C = 1) and addition (C = 2,3).

Contrast Correlations.

Correlations between SCORE3 main effect contrast scores and reference tests are summarized in Table 3.8.

Correlations between the N contrast and reference tests were previously discussed (see Table 3.1). To recapitulate, mean SCORE3 scores correlated significantly with ten spatial or Gf reference tests, and no verbal tests. Individual construction linear and quadratic contrast scores showed strong differential correlations with several reference spatial tests. The largest differentials were with Form Board and WAIS Object Assembly. The linear contrast scores correlated .53 and .56 with Form Board and Object Assembly, while the corresponding quadratic contrast correlations were -.52 and -.51. Thus, high scores on these tests are reflected in a linear decline in correctness over the construction facet.

The only significant correlation for the rotation contrast scores was a correlation of .44 between the linear contrast and Figures. The correlation with Cards was similar but nonsignificant (r=.39). The rotation quadratic component correlations were not significant.

While the stimulus complexity constrasts had no significant correlations with reference tests, the product complexity contrast correlated negatively with two WAIS verbal subtests and the Verbal Scale Score (r=-.43). Thus, large decrements in performance over the product complexity facet are more strongly associated with verbal ability than spatial ability.

Preliminary Analysis of Confidence

Although the confidence data were used primarily to help "clean up" the correctness index, they are of some interest in their own right. In general, the analysis of confidence mirrored

Table 3.8

Correlations between SCORE3 Main Effect Contrast Scores and Reference Tests (N=30)

N		C Linear		C Quadratic		R Linear		R Quadratic
IQ	51	Form Board	53	IQ	-45	Cards	(39)	
Paper Folding	68	Copying	42	Form Board	-52	Figures	44	
Form Board	48	Figures	52	Mem for Designs-Q	-42			
Surface Develop.	65	Mem for Designs-Q	58	Total Draw	-56			
Cards	43	Total Draw	50	Identical Pic.	-48			
Figures	50	Identical Pic.	44	Object Assembly	-51			
Mem for Designs	52	Object Assembly	56	WAIS Perform. S.	-49			
Mem for Designs-Q	68	WAIS Perform. S.	51					
Total Draw	69							
Vis Num Span	43							
Letter Series	52							
Picture Arrange.	48							
WAIS Perform. S.	51							

S Linear	S Quadratic		P	
	Digit Span-F	-44	Comprehension	-45
			Similarities	-45
			WAIS Verbal S.	-43

Note. Decimals and correlations less than .42 omitted.

the analysis of a composite index of difficulty (time minus correctness) that is reported below.

The analysis of variance on these scores is reported in Lohman (1979b). Confidence decreased significantly over all facets except Match. For rotation and stimulus complexity, the decline was linear (the quadratic contrasts were not significant). For construction, the decline was greater between no addition and one addition than between one and two piece additions.

There were a number of significant interactions. However, all were congruent with other analyses, particularly ZDIFF (see below) and total time. Therefore, they will not be discussed here. The important finding here was that immediately obtained self report surity ratings closely paralleled the more "objective" latency and correctness measures. Thus, self reports may have considerably more validity than is commonly assumed (e.g., Nisbitt and Wilson, 1977), especially when obtained as in this experiment.

Comparison of SCORE3 and Total Time Analyses.

The separate analyses on total time and SCORE3 clearly reveal that time and error rate are not interchangeable aspects of performance.

Since the data were collapsed over match and type of addition for the correctness analyses, the analyses of variance on total time and SCORE3 are not totally comparable. There were, however, two important differences between them.

1. Error rate was the same at levels two and three of the construction facet, while total time increased linearly over the three levels of construction.
2. Interactions were fewer and smaller in the SCORE3 analysis than in the total time analysis.

The most striking differences between the two analyses, however, were in the correlations between individual contrast scores and reference tests. These differences are summarized below for each contrast.

N Contrasts.

Mean SCORE3 correlated significantly with ten spatial or Gf tests. Mean time, on the other hand, correlated significantly with one verbal reasoning test. Correlations with several other verbal tests approached significance, while correlations with spatial tests centered near zero.

Construction.

For SCORE3, a linear decline was associated with high scores and a quadratic decline with low scores on Form Board and Object assembly. For total time, a linear increase was associated with verbal ability. The quadratic contrast scores for total time correlated positively with Marks Vividness of Visual Imagery Questionaire and negatively with Hidden Figures.

Rotation.

The linear decline in SCORE3 over the rotation facet was correlated significantly with only one reference test: Figures. The correlation with the Cards test was similar but not significant. For total time, the magnitude of the linear increase was correlated with number of errors on the Cards test.

Stimulus Complexity.

Neither the linear nor quadratic SCORE3 stimulus complexity contrasts correlated with any of the reference variables. Both total time contrasts, on the other hand, correlated strongly with verbal tests.

Product Complexity.

For SCORE3, large decrements in performance over this facet were associated with verbal ability. For total time, large increments were associated with general ability.

Match and Type of Addition.

Since data were collapsed over these facets in the SCORE3 analysis, only total time contrast scores were obtained. For the

match facet, faster responses to positive test stimuli were associated with high scores on Form Board and Picture Completion. Faster responses to negative test stimuli were associated with more errors on the Cards test and higher scores on Visual Number Span.

Taking much longer on left addition than right additions correlated highly with Object Assembly and moderately with several verbal tests.

Summary of the Separate Analyses.

Correlations between the two N contrasts and the reference tests address the construct validity of the test in global terms. They both answer the question, "What, in general, is measured by this task?" For SCORE3, the answer is clearly "spatial ability." For total time, it is a more tentative "verbal ability, if anything."

Given these global assessments, the contrast correlations within each facet look at the relationships between the pattern of scores over that facet and the reference tests.

In general, the SCORE3 contrast correlations supported the construct validity of the task, while the total time correlations revealed more about the strategies or processes subjects used to solve particular item types. These correlations have already been discussed in some detail for the stimulus complexity, product complexity, match and type of addition facets. Construction and rotation, however, have been examined only briefly. They are the two "process" facets and thus merit closer scrutiny.

The differential correlations between the construction contrasts, and Form Board and Object Assembly support the hypothesis that mental construction is an identifiable spatial skill. Total time contrasts for this facet reveal that those who took much longer to solve the two addition items than would be predicted by their "basal" time on no construction items were probably reverting to verbal strategies. Even more interesting was the finding that those who reported strong visual imagery

showed smaller increases in time between two and three piece additions than between no addition and one piece addition. On the other hand, those whose times increased the most between two and three piece additions may have been using a more analytic spatial strategy.

For the rotation facet, the correlation between the linear contrast scores and the two rotation reference tests, Cards and Figures, support the construct validity of the facet. Large increases in time over the facet were associated with number of errors on the Cards test. Examination of component times and their correlations suggests that those who made many errors on the Cards test may have been using a verbal strategy to solve some or all of the items. Thus, once again time reveals something about process while correctness contrasts address the construct validity of the facet.

IV. Combined Analyses on Total Time and Correctness

The Advantages of Combined Analyses.

While separate univariate analysis are useful in providing an overview of the major results, they have a number of disadvantages. First, it is difficult to coordinate the results of several analyses on correlated measures. More important, however, is the fact that such analyses may be misleading. This is more a product of the experimental dependence of time and correctness than of their statistical dependence.

The interaction between rotation and product complexity provides an example of the ambiguities that arise when time and correctness are analyzed separately. The interaction was highly significant for total time, marginally significant for SCORE3 (correctness adjusted for confidence and match time), and nonsignificant for SCORE (correctness). Plots of the cell means for each of the three dependent variables are shown in the top row of Figure 3.3. The plot for total time reveals an almost perfect linear relationship between time and rotation for low complexity constructed stimuli. However, it appears that, for constructed stimuli of high complexity, 180 degree rotations may

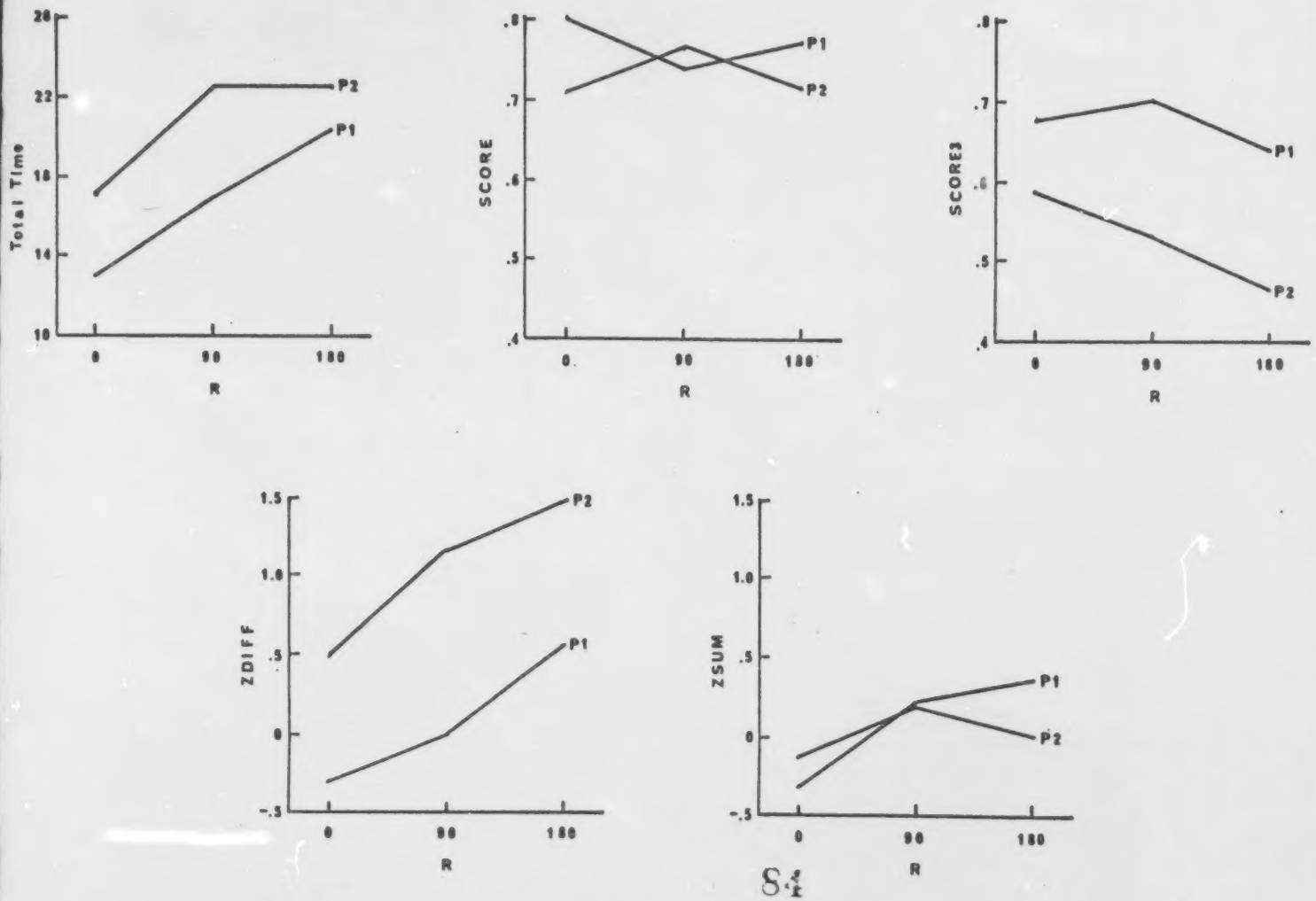

Figure 3.3. Plots of the rotation by product complexity interaction for total time, SCORE, SCORE3, ZSUM, and ZDIFF.

be performed as quickly as 90 degree rotations.

While the plots for SCORE do not shed any light on the problem, the plot for SCORE3 is suggestive. The latter plot shows a linear decrease in correctness for high product complexity items. Note also that probability correct (by SCORE3) for high product complexity is almost at .5 for 90 degree rotations and then below it for 180 degree rotations. This suggests that a 90 degree rotation of a complex constructed stimulus was about all the average subject could manage. 180 degree rotations of such stimuli were just too difficult, and so subjects gave up and resorted to guessing.

Sum and Difference Score Analyses

While such coordinate examination of error and latency data is useful, there is a better way to examine their joint and independent effects. Time and correctness (by SCORE3) were first independently standardized to zero mean and unit variance. Two scores were then derived by taking the sum and the difference of these standard scores. ZSUM was the sum of the standard scores for total time and correctness, and ZDIFF time minus correctness. ZDIFF contrasts fast-accurate (low negative) and slow-inaccurate (high positive) performance, while ZSUM contrasts fast-inaccurate (low negative) and slow-accurate (high-positive) performance. Variations in facet difficulty or complexity should be reflected in ZDIFF, i.e., fast, accurate responses to easy items and slow, inaccurate responses to difficult items. ZSUM, on the other hand reflects variations in speed-accuracy tradeoff.

Plots of ZDIFF and ZSUM for the Rotation by Product Complexity interaction are shown in the bottom row of Figure 3.3. Although the quadradic component of the interaction is still significant in this picture, the slopes of the lines at the two levels of product complexity are more nearly parallel. Further, the ZSUM plot reveals a large shift in speed-accuracy tradeoff for high product complexity items. Subjects may have given up too soon or mayhave been unable to perform these difficult items. Later results favor the latter interpretation.

Since these two derived scores are statistically independent, each may be analyzed separately and the results then compared unambiguously.

Analyses of variance on ZSUM and ZDIFF are presented in Tables 3.9 and 3.10. Estimated variance components are reported elsewhere (see Lohman, 1979b). In the ZDIFF analysis, design facets accounted for 52.3 percent of the variation in scores, and subjects and their interactions with design facets accounted for 44.0 percent. Thus, item difficulty was a product of both design facets and subject interactions. For ZSUM, on the other hand, the design facets accounted for only 9.4 percent of the variation in the data. Subjects and their interactions with design vectors accounted for 84.9 percent. Thus, the speed-accuracy tradeoff was only slightly influenced by design facets; rather, it was largely a function of subjects and how they responded to particular item types. These relationships are shown graphically in Figure 3.4.

Interpretation of the Sum and Difference Score Analyses

N Contrast.

This contrast reflects variability in the subject means about the grand mean. For ZDIFF, the correlations with spatial reference tests were mostly negative but nonsignificant. The only significant correlation was -.50 with WAIS Picture Arrangement. The correlations with verbal tests were zero or slightly positive. Thus, except for the WAIS Picture Arrangement subtest, fast-accurate performance on this task was only weakly related to fast-accurate performance on other spatial tasks.

The N contrast for ZSUM, on the other hand, had a number of large positive correlations with reference tests: Total Draw (.66), Paper Folding (.55), Letter Series (.55), Visual Number Span (.54), Terman Concept Mastery (.52), Memory for Designs-Quality (.51), WAIS Information (.47), and Verbal Scale Score (.44). Although verbal tests are represented on this list, the higher correlations are associated with complex spatial and reasoning tests. The pattern is congruent with Cattell's general

Table 3.9

ZSUM Analysis of Variance

Source	df	MS	F	Source	df	MS	F
Mean	1	.00	-	N	29	32.19	-
Construction	(2)			NC	(58)		
Linear	1	39.95	8.36*	Linear	29	4.78	9.53*
Quadratic	1	22.70	18.66*	Quadratic	29	.85	1.70
Rotation	(2)			NR	(58)		
Linear	1	15.90	7.17	Linear	29	2.22	4.43*
Quadratic	1	11.35	18.02*	Quadratic	29	.63	1.26
Stimulus Complexity	(2)			NS	(58)		
Linear	1	15.90	8.94*	Linear	29	1.78	3.55*
Quadratic	1	.74	1.88	Quadratic	29	.39	.79
Product Complexity	1	2.09	.95	NP	29	2.21	4.41*
CR	4	7.69	11.10*	NCR	116	.69	1.38
CS	4	2.13	4.25*	NCS	116	.64	1.28
KP	1	16.43	15.31*	NKP	29	1.07	2.14*
RS	4	.97	1.84	NRS	116	.52	1.05
RP	2	5.24	7.45*	NRP	58	.70	1.40
SP	2	3.54	3.56	NSP	58	.99	1.98*
CRS	8	1.52	2.73*	NCRS	232	.56	1.11
KRP	2	.14	.31	NKRP	58	.47	.93
KSP	2	1.58	3.19	NKSP	58	.50	.99
RSP	4	5.06	6.09*	NRSP	116	.83	1.66*
KRSP	4	1.32	1.63	NKRSP	116	.81	1.61*
				Residual	270	.50	-

* = p less than .01

Table 3.10

ZDIFF Analysis of Variance

Source	df	MS	F	Source	df	MS	F
Mean	1	.00	-	N	29	22.77	-
Construction	(2)			NC	(58)		
Linear	1	1099.89	305.53*	Linear	29	3.60	6.59*
Quadratic	1	94.73	119.76*	Quadratic	29	.79	1.45
Rotation	(2)			NR	(58)		
Linear	1	255.79	122.80*	Linear	29	2.08	3.81*
Quadratic	1	3.82	4.43	Quadratic	29	.86	1.58
Stimulus Complexity	(2)			NS	(58)		
Linear	1	153.91	89.01*	Linear	29	1.73	3.17*
Quadratic	1	.92	1.79	Quadratic	29	.51	.94
Product Complexity	1	243.85	93.21*	NP	29	2.62	4.79*
CR	4	1.24	2.11	NCR	116	.59	1.07
CS	4	11.16	18.19*	NCS	116	.61	1.12
KP	1	8.14	6.71	NKP	29	1.21	2.22*
RS	4	.97	1.65	NRS	116	.58	1.07
RP	2	5.74	6.88*	NRP	58	.84	1.53
SP	2	8.06	8.96*	NSP	58	.90	1.64*
CRS	8	2.73	3.95*	NCRS	232	.69	1.27
KRP	2	3.55	5.54*	NKRP	58	.64	1.17
KSP	2	25.64	31.93*	NKSP	58	.80	1.47
RSP	4	2.36	3.09	NRSP	116	.76	1.40
KRSP	4	23.79	27.27*	NKRSP	116	.87	1.60*
				Residual	270	.55	-

* = p less than .01

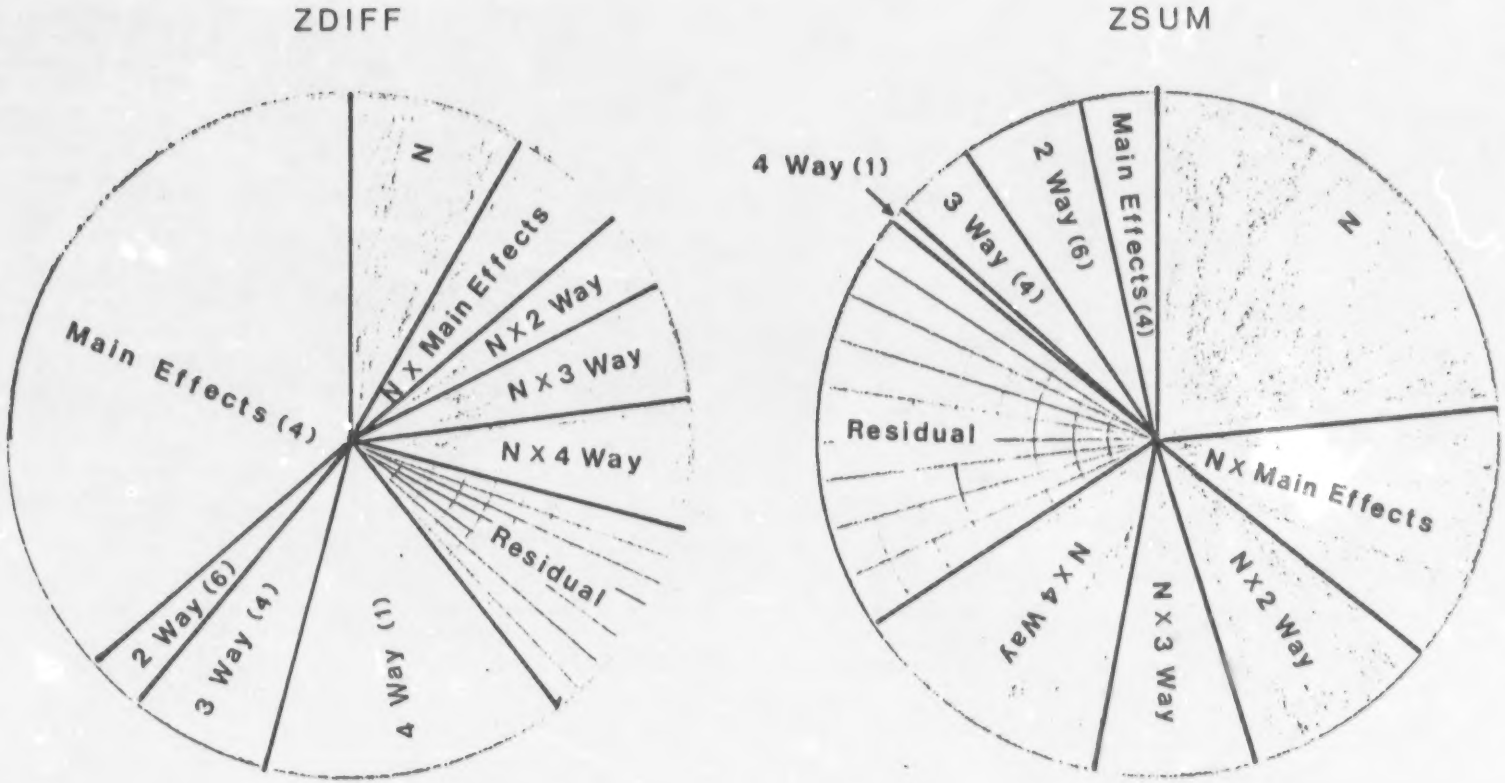

Figure 3.4. Partitioning of variance for ZDIFF and ZSUM.

fluid ability factor (Gf). Thus, the general tendency to opt for accuracy over speed was associated with Gf or general ability (G).

Construction.

Construction was the most powerful design facet. Both the linear and quadratic effects were highly significant for ZDIFF. As can be seen in Figure 3.5, the quadratic effect reflects a larger increase in difficulty between no construction (C=1) and one piece construction (C=2), than between one and two piece additions. Subject interactions with the linear term were also highly significant, reflecting large individual differences in slopes. However, there were no significant correlations between the construction linear contrast and reference constructs. Individual scores for the quadratic construction contrast, however, correlated .51, .43, .54 and .45 with Form Board, total number of correct draw trials, WAIS Object Assembly, and WAIS Performance Scale Score, respectively. Thus those subjects who made relatively more errors and had relatively longer latencies on level 3 construction items than on level 2 items did better on the draw trials and on these reference tests. In other words, subjects who "topped out" at level 2 construction items tended to have lower scores on these three variables.

A plot of the ZSUM means over the construction facet is also shown in Figure 3.5. Since subject interactions with both the linear and quadratic components were large, the plot is not representative of individual data. The most striking relationships here were the high positive correlations between the linear contrast scores and verbal and memory span reference constructs, and high negative correlations between quadratic contrast scores and spatial reference constructs. Further, the two contrasts had opposite correlations with Total Draw, .60 for the linear contrast and -.63 for the quadratic contrast. The quadratic contrast correlations indicate that those who showed a much larger shift in speed-accuracy tradeoff between one and two piece additions than between zero and one piece additions

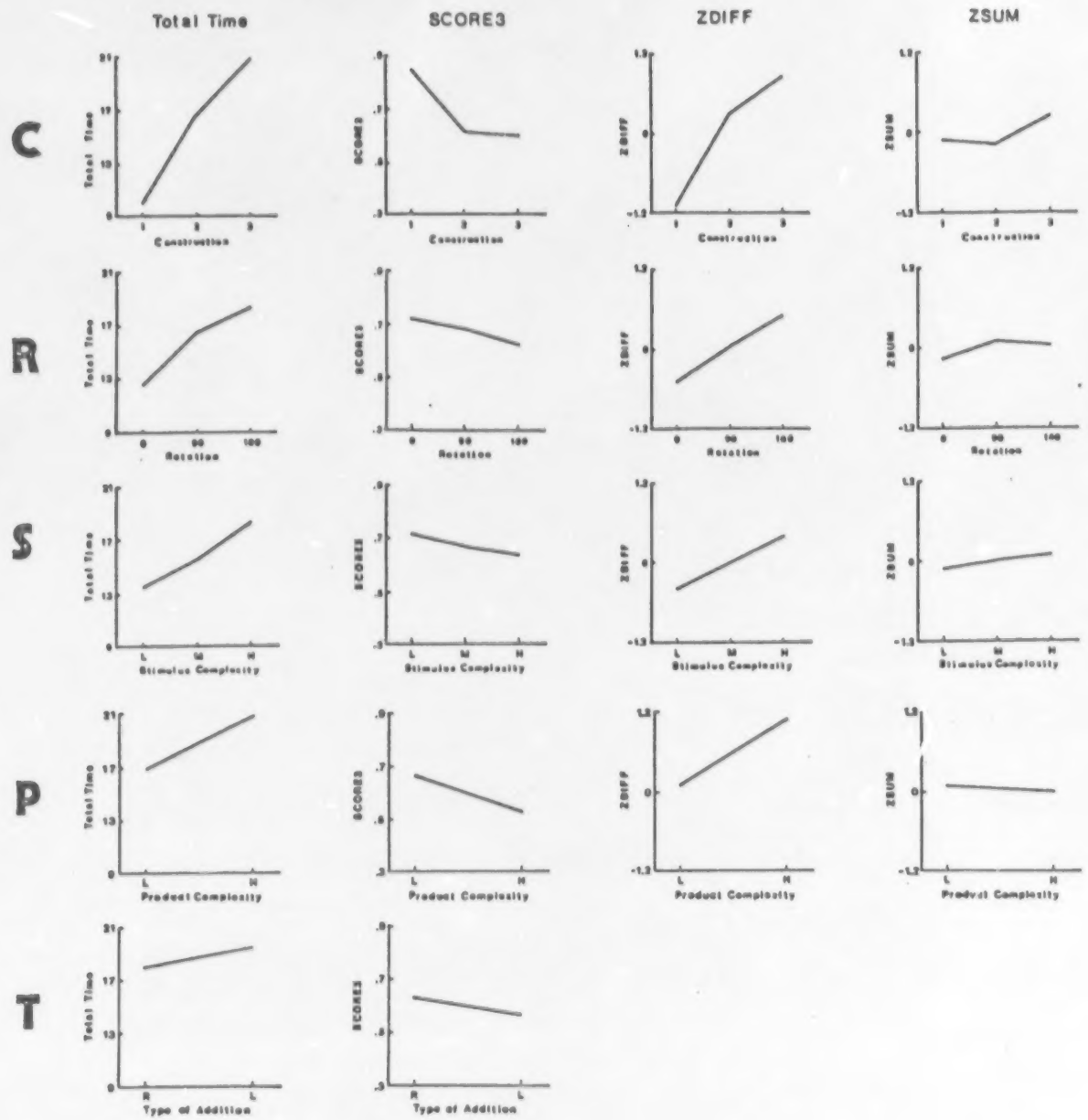

Figure 3.5. Plots of main effect cell means for total time, SCORE3, ZDIFF, and ZSUM.

performed poorly on the reference spatial tests and the draw trials. On the other hand, large shifts between no construction and two piece additions were associated with verbal and verbal memory span tests.

Rotation.

For ZDIFF, the linear contrast was highly significant, and the quadratic contrast only weakly significant. Figure 3.5 reveals the nearly perfect linear relationship between ZDIFF and rotation. Similarly, subject interactions with the linear component were strong, while the subjects by rotation quadratic term was only marginally significant. There were no significant correlations between either contrast and the reference tests.

Again the ZSUM contrasts had several noteworthy correlations. The linear and quadratic contrasts corelated .45 and -.22 with number of errors on the Cards tests. Corresponding correlations with the Total Draw score were .44 and -.02. Thus, those who shifted to slower, more accurate responding for 180 degree rotations did better when they had to draw the answer, but made more errors on the multiple choice Cards test.

Stimulus Complexity.

For ZDIFF, only the linear contrast and subjects by linear contrast were significant. The straight line plot is shown in Figure 3.5. It is noteworthy that the analyses of total time suggested significant nonlinear trends for stimulus complexity and rotation. The plots became almost perfectly linear only when latency and correctness were analyzed jointly. The linear contrast had only two significant correlations: .53 with WAIS Verbal Scale Score, and .67 with WAIS Digit Span Backwards. The quadradic component correlated .48 with Visual Number Span and .55 with WAIS Digit Span Forward. Together, these correlations suggest that those who showed the strongest effects for stimulus complexity may have been using verbal strategies to encode and remember the stimuli.

The plot for ZSUM on stimulus complexity is shown in Figure

3.5. Again, only the linear and subjects by linear contrasts were significant. As with ZDIFF, correlations with reference tests again suggest a connection with verbal ability. Both linear and quadratic contrast scores correlated .45 with WAIS IQ. Only the linear contrast had other significant correlations. Four were with verbal tests (Word Beginnings and Endings, .50; WAIS Comprehension, .43, Vocabulary, .47, and Verbal Scale Score, .48) while two were with spatial tests (Cards Wrong, .50; and WAIS Object Assembly, .44). However, Cards Wrong is probably a better index of verbal strategies than spatial ability. Thus, large shifts toward slower, more accurate performance over the rotation facet were most strongly related to verbal ability and verbal strategies.

Product Complexity.

As shown in Table 3.10, product complexity and subjects by product complexity were highly significant in ZDIFF analysis. Both errors and latency were greater for high complexity constructed stimuli.

Contrast scores for product complexity were the best predictor of performance on the reference tests. High correlations with both verbal and spatial tests implicate general ability. In particular, the performance of high ability students was influenced by the complexity of the to-be-constructed stimulus. This is most easily explained by assuming that high ability students were able to combine the stimuli mentally as directed, (using verbal strategies spatial strategies or both) while low ability students were not able to generate a clear image of the constructed stimulus.

This result is even more interesting in the light of the ZSUM analysis. While the subjects by product complexity interaction was significant for ZSUM, the main effect was not. However, none of the contrast correlations were significant. Hence, the product complexity effect is not a reflection of systematic changes in speed-accuracy tradeoff.

Defense of Sum and Difference Scores

The propriety of any index depends on the question that is asked. Thus, questions about the relationships between latency or correctness and the design facets or reference tests are best addressed by separate analyses on the two measures.

But, for example, reading and mathematics achievement test scores may be usefully combined to represent two new variables, namely, general scholastic achievement, and differential achievement in reading and mathematics. Verbal and spatial scores also may be summed to represent general abilty and subtracted to provide a special ability contrast. Similarly, latency and correctness can be combined to form new indices. The major difference between the separate scores and the two combination indices lies in their interpretation. Further, unlike the other examples cited above, latency and correctness are not only statistically, but also experimentally dependent. This also affects their interpretation.

The difference score (time minus correctness) is best interpreted as a composite index of difficulty, while the sum represents both the speed-accuracy tradeoff and the independent contributions of the two variables.

If latency and correctness have opposite patterns over a facet (e.g., latency increases linearly while correctness decreases linearly), then the means analysis on the difference score will show a larger effect for the facet than either of the separate analyses. Similarly, the sum score will be nonsignificant over the facet, reflecting the fact that the patterns are the same, but opposite, for the two variables.

For the correlational analysis, if both latency and correctness correlate positively with a given test, then the sum score correlation will be higher, and the difference score correlation zero. If latency correlates with verbal tests and correctness with spatial tests, then the sum will correlate with both and the difference with neither cluster of tests. In such cases, separate analyses on latency and correctness provide greater insight.

However, there are times when the sum and difference scores provide unique information. Several puzzling interactions appearing in the separate analyses were eliminated or attenuated in the combined analyses (e.g., see Figure 3.3). Similarly, both rotation and stimulus complexity showed nonlinear trends in the total time analysis, but were linear in the ZDIFF analysis. Thus, "difficulty" increased linearly over these facets, while time and correctness had nonlinear, but opposite, quadratic effects.

Similarly, correlations between the sum and difference scores contrasts were often quite unlike the contrast correlations in the separate analyses. For example, in the analysis on construction component time and construction SCORE3 (see below), the type of addition contrast for component time correlated significantly with seven verbal tests (r = .41 to .56) and Object Assembly (r = .45). The SCORE3 type of addition contrast correlated with Marks Vividness of Visual Imagery Questionaire (r = .49) and negatively with Visual Number Span (r = -.45). However, the sum score correlated with Form Board (r = .43) and Object Assembly (r = .46), while the difference score correlated negatively with Marks Questionaire (r = -.56) and positively with Visual Number Span. Thus, the verbal tests dropped from the picture and the two tests most similar to the experimental construction task (Form Board and Object Assembly) became the only two tests correlated with the sum score. The sum score correlations occurred because Form Board and Object Assembly were positively correlated with both the time contrast (r = .32 and .45) and the SCORE3 contrast (r = .32 and .28), although the three smaller correlations were not significant. On the other hand, the SCORE3 type of addition contrast had zero or slightly negative correlations with the verbal tests. Thus, when SCORE3 and time were added together, the two tests positively correlated with both the time and correctness contrasts emerged with larger correlations, while the verbal tests correlations became nonsignificant.

Summary of the Sum and Difference Score Analyses

The analyses of variance on ZSUM and ZDIFF differed most significantly in the way score variation was accounted for by facets and subject interactions with facets. In particular, the ZDIFF analysis revealed that item difficulty (on a scale of fast-accurate to slow-inaccurate) was a function of both facets and subject interactions with facets. The speed-accuracy tradeoff represented by ZSUM, on the other hand, was only slightly influenced by design facets. Rather, variation in these scores was largely a function of subject interactions with design facets.

Most of the relationships between ZSUM and ZDIFF contrast scores and reference tests are congruent with the previous analyses on total time and SCORE3. There are important differences, however, and these are noted below.

N Contrast.

For ZDIFF, generally fast and accurate performance on this task was weakly related to similar performance on other spatial tasks, and strongly indicative of high scores on the WAIS Picture Arrangement subtest. There was no relationship, or, if anything, an inverse relationship between ZDIFF and performance on Verbal tests.

For ZSUM, the general tendency to opt for accuracy over speed was associated with Gf.

Construction.

As in the SCORE3 analysis, nonlinear increases in ZDIFF were positively related to Form Board and Object Assembly. Those who did not take more time and make more errors on two piece additions than on one piece additions performed poorly on two reference construction tests.

One explanation for this phenomenon would be that those who performed poorly on these construction tasks did not fully appreciate the increased difficulty of the two piece additions, and thus did not really try harder on the more difficult

additions. However, the ZSUM contrast correlations argue against this hypothesis. Large shifts in speed-accuracy between levels two and three on the construction facet were also related to poor performance on the reference tests. Therefore, a better explanation of the ZDIFF quadratic correlations is that those who did not show an increase in error rate and time when shifting from one to two piece additions had reached their limit on the one piece additions.

Finally, the ZSUM linear correlations related large shifts in speed-accuracy trade off between no construction and two additions to verbal ability.

Rotation.

Only the ZSUM linear contrast had significant correlations for the rotation facet. Large shifts toward slower, more accurate performance were related to number of errors on the Cards test and Visual Number Span.

Stimulus Complexity.

The ZSUM and ZDIFF analyses indicated that those who showed the largest effects for this facet may have been using verbal strategies to encode and remember the stimuli. The major difference between these correlations and those for the same contrast in the total time analysis was the focus on verbal short term memory tests in the ZDIFF analysis.

Product Complexity.

The most significant contribution of the ZSUM, ZDIFF analyses here was the finding that the product complexity effect was not a function of systematic changes in speed-accuracy trade-off.

V. Component Time Analyses

At least two component times (encode, match) and as many as four component times (encode, add, rotate, match) were collected for each item. Encoding and match times were available for all

216 items, construction and rotation times for 144 items. Since construction always preceded rotation, rotation component times for the 96 items that also required construction were not experimentally independent of the construction operations. Rotation time on these items depended not only on the success of the rotation, but also on the success of the preceding addition. Similarly, only 24 match times were not preceded by construction or rotation. Therefore, the analyses of the rotation and match component times were necessarily more limited than those on encoding and construction.

Encoding

Means Analysis.

Stimulus complexity and practice (or block) were the only two factors represented in the design likely to affect encoding. The analysis of variance of encoding time by these factors is reported in Table 3.11.

Only the linear components of both stimulus complexity and block were significant. Cell means were 4.02, 5.41, and 6.92 sec for low, medium, and high complexity stimuli, respectively. Cell means for the four blocks were 6.71, 5.33, and 5.25, and 4.49 sec.

There were significant individual differences in both linear contrasts and in the nonlinear block contrast. Subjects also interacted significantly with the linear by linear contrast.

Contrast Correlations

N Contrast. The first important finding is a negative one: there were no relationships between average encoding time and any of the reference tests or total scores. The correlation between mean encoding time and IQ was -.03. The only significant correlation (r=.48) was with number of marks made on the Form Board test, a strategy index. Those who made more marks on the stimuli in the Form Board test spent more time studying the stimuli in this experiment.

Stimulus Complexity. The stimulus complexity linear

Table 3.11

Analysis of Variance for Encoding Component Time

Source	df	MS	F	Source	df	MS	F[a]
Mean	1	10683.44	176.08	N	29	60.67	-
Stimulus Complexity	(2)			NS	(58)		
Linear	1	503.07	69.90*	Linear	29	7.20	11.80*
Quadratic	1	.32	.48	Quadratic	29	.68	1.11
Block	(3)			NB	(87)		
Linear	1	203.11	18.33*	Linear	29	11.08	18.16*
Nonlinear	2	26.46	2.05	Nonlinear	58	6.45	10.57*
SB	(6)			NSB	(174)		
Lin x Lin	1	2.81	1.85	Lin x Lin	29	1.52	2.50*
Qd x Nln	5	.82	1.34	Qd x Nln	145	.61	-

[a]MS(e) estimated by the NS(quadratic)B(nonlinear) interaction.

*p less than .01

contrast had significant correlations with WAIS Digit Span (r=.47), Vocabulary (r=.44), Verbal Scale Score (.49), and thus IQ (.43). The correlation with WAIS Performance Scale Score was a nonsignificant .27. Only one spatial test, Paper Folding, correlated significantly with the contrast (r=.47). There was also an interesting differential set of correlations with the two correctness indices: r=.43 with total SCORE, but only .14 with total SCORE3.

The quadradic component had even higher correlations with verbal tests. Correlations with all WAIS verbal subtests except Arithmetic were significant, and the correlation with Verbal Scale Score was .58. Other significant correlations were total correct Draw trials (.52), Visual Number Span (.44), Number Comparison (.43), and Camouflaged Words (.54).

These quadratic component correlations mean that those who showed greater increases in encoding time between medium and high complexity than between low and medium complexity tended to score higher on verbal tests than those whose times increased linearly or by the reverse pattern.

Strategy reports provide a good explanation of this phenomenon. Some subjects reported that they tried to remember complex stimuli by seeing them as composites of simpler forms such as triangles and squares. Both additive and subtractive versions of this strategy were reported , i.e., "a triangle on top of a square," and "a square with a triangle cut out of the top."

Subjects using this strategy would be much faster on low and medium complexity stimuli (3 - 4 points) since most of these stimuli can be immediately labelled, even when irregular (e.g. "a triangle leaning to the right"). Most high complexity stimuli were not so familiar, and so the process of decomposing them into simpler shapes or looking for some concrete association would have taken longer.

The Construction Component

Of the total 216 items, 144 had a construction component.

Construction component times were relatively "clean" since rotation, if required, always followed construction. Data were collapsed over the rotation facet, and mean time and number correct within each of these 24 cells were then used in the analyses. Note that the construction facet has only two levels here.

Separate analyses of variance were performed on time and correctness (by SCORE3) and are reported in Lohman (1979b). The major differences between the two analyses were:

1. Construction (K) was a highly significant facet in the time analysis (F=118.79), but not significant in the SCORE3 analysis (F=.44). Thus, subjects took much longer to add two stimuli to a given base than to add one. However, they did not make relatively more errors on the two piece additions.
2. The KP, ST, and KSTP interactions were highly significant in the time analysis but not significant in the SCORE3 analysis.
3. In the KS and SP interactions, the linear component of the contrast was highly significant in one analysis and the quadratic component significant in the other analysis.

As in previous analyses, time and correctness were analyzed jointly by computing the sum and difference of the independently normalized scores. These variables were called CSUM and CDIFF. Analyses of variance on each are reported in Tables 3.12 and 3.13. Cell means are reported elsewhere (see Lohman, 1979b).

In the CDIFF analysis, all main effects and most facet interactions were significant, while only the NK, NP, NKP, and NKTP subject-facet interactions were significant. On the other hand, construction (K) and product complexity were the only significant main effects in the CSUM analysis. Most facet interactions were significant, but generally smaller than their counterparts in the CDIFF analysis. Large subject-facet interactions were obtained for K, S linear, and P.

Table 3.12

CSUM Analysis of Variance

Source	df	MS	F	Source	df	MS	F
Mean	1	.000	-	N	29	19.674	-
Construction (K)	1	122.776	81.69*	NK	29	1.503	2.90*
Stimulus Complexity	(2)			NS	(58)		
Linear	1	.994	.78	Linear	29	1.278	2.46*
Quadratic	1	.888	2.04	Quadratic	29	.435	.84
Type of Addition	1	.376	.47	NT	29	.801	1.54
Product Complexity	1	.248	.12	NP	29	2.023	3.90*
KS	(2)			NKS	(58)		
Linear	1	7.057	15.86*	Linear	29	.445	.86
Quadratic	1	.000	.00	Quadratic	29	.669	1.29
KT	1	.143	.15	NKT	29	.938	1.81*
KP	1	16.119	17.43*	NKP	29	.925	1.78
ST	(2)			NST	(58)		
Linear	1	.571	1.05	Linear	29	.545	1.05
Quadratic	1	.721	.93	Quadratic	29	.773	1.49
SP	(2)			NSP	(58)		
Linear	1	13.913	16.99*	Linear	29	.819	1.58
Quadratic	1	4.152	6.19	Quadratic	29	.671	1.29
TP	1	1.860	2.99	NTP	29	.622	1.20
KST	(2)			NKST	(58)		
Linear	1	.404	1.01	Linear	29	.400	.77
Quadratic	1	4.728	11.07*	Quadratic	29	.427	.82
KSP	(2)			NKSP	(58)		
Linear	1	3.347	5.86	Linear	29	.571	1.10
Quadratic	1	2.902	3.50	Quadratic	29	.828	1.60
KTP	1	7.864	7.49	NKTP	29	1.050	2.02*
STP	(2)			NSTP	(58)		
Linear	1	.758	1.19	Linear	29	.639	1.23
Quadratic	1	3.675	4.65	Quadratic	29	.790	1.52
KSTP	(2)			NKSTP	(58)		
Linear	1	4.586	9.38*	Linear	29	.489	-
Quadratic	1	1.236	2.25	Quadratic	29	.549	-

[a]MS(e) estimated by NKSTP interaction.

*p less than .01

Table 3.13

CDIFF Analysis of Variance

Source	df	MS	F	Source	df	MS	F[a]
Mean	1	.000	-	N	29	11.188	-
Construction (K)	1	140.747	97.81*	NC	29	1.439	2.44*
Stimulus Complexity	(2)			NS	(58)		
Linear	1	70.411	96.98*	Linear	29	.726	1.23
Quadratic	1	5.599	11.13*	Quadratic	29	.503	.85
Type of Addition	1	42.912	53.91*	NT	29	.796	1.35
Product Complexity	1	253.171	109.17*	NP	29	2.319	3.94*
KS	(2)			NKS	(58)		
Linear	1	8.619	15.23*	Linear	29	.566	.96
Quadratic	1	6.459	16.69*	Quadratic	29	.387	.66
KT	1	33.391	51.45*	NKT	29	.649	1.10
KP	1	8.619	8.16*	NKP	29	1.056	1.79
ST	(2)			NST	(58)		
Linear	1	17.511	26.57*	Linear	29	.659	1.12
Quadratic	1	.192	.29	Quadratic	29	.666	1.13
SP	(2)			NSP	(58)		
Linear	1	16.251	23.52*	Linear	29	.691	1.17
Quadratic	1	4.401	5.65	Quadratic	29	.779	1.32
TP	1	28.227	38.61*	NTP	29	.731	1.24
KST	(2)			NKST	(58)		
Linear	1	13.187	24.11*	Linear	29	.547	.93
Quadratic	1	.499	.82	Quadratic	29	.612	1.04
KSP	(2)			NKSP	(58)		
Linear	1	7.203	13.72*	Linear	29	.525	.89
Quadratic	1	25.007	32.23*	Quadratic	29	.776	1.32
KTP	1	25.560	21.48*	NKTP	29	1.190	2.02
STP	(2)			NSTP	(58)		
Linear	1	1.532	2.02	Linear	29	.760	1.29
Quadratic	1	5.217	7.46	Quadratic	29	.699	1.19
KSTP	(2)			NKSTP	(58)		
Linear	1	.001	.00	Linear	29	.672	-
Quadratic	1	7.530	14.88*	Quadratic	29	.506	-

[a]MS(e) estimated by NKSTP interaction.

*p less than .01

Construction Facet.

As previously noted, K was a powerful facet in the component time analysis, but not significant in the SCORE3 analysis of variance. This emerged as a large main effect in both the CSUM and CDIFF analyses. The latter reflects the fact that two piece additions were more difficult than one piece additions. The CSUM main effect reflects the large changes in speed-accuracy tradeoff over the facet. Subjects maintained a constant error rate by taking an average of 4.56 sec longer to solve the two piece additions.

Individual contrast scores on CDIFF correlated significantly with Cards Wrong (.48), total correct Draw trials (.49), Necessary Arithmetic Operations (.43), Terman Concept Mastery (.55), WAIS Information (48), Vocabulary (46), and Verbal Scale Score (.49). Thus high verbals were most drastically affected by the shift from one to two piece additions.

The CSUM contrast scores had only one significant correlation: .56 with Visual Number Span Longest Correct. Perhaps those good at remembering numbers tend to be poor at remembering figures.

Stimulus Complexity.

Both linear and quadratic stimulus complexity contrasts were significant in the CDIFF analysis. The latter reflects a greater increase in difficulty between medium and high than between low and medium complexity. Neither contrast was significant in the CSUM analysis.

Individual CDIFF linear contrast scores correlated significantly with a cluster of tests similar to those found on factors called Reasoning in the factor analytic literature. The tests and their correlations were: Necessary Arithmetic Operations (.43), WAIS Arithmetic (.50), Digit Span Backwards (.47), and Verbal Scale Score (.45). On the other hand, the CDIFF quadratic contrast scores correlated most strongly with spatial reference tests. Seven of ten significant correlations were with spatial tests. Thus, for subjects with high spatial

scores, additions involving medium complexity stimuli were only slightly more difficult than those involving low complexity stimuli. Additions of high complexity stimuli were much more difficult.

For CSUM, the linear contrast scores correlated with Word Beginnings and Endings (.55), Uses for Things (.51), and WAIS Vocabulary (.44). Once again, large changes in speed-accuracy tradeoff over the stimulus complexity factor are associated with verbal ability, this time with a cluster more properly called Word Fluency or Verbal Productive Thinking.

Type of Addition.

While the main effect indicates that adding to the left was generally more difficult than adding to the right, the KT interaction reveals that this difference held only for two piece additions. Contrast scores on CDIFF for this factor correlated -.56 with Marks Vividness of Visual Imagery Questionaire and .56 with Visual Number Span. Thus, those who had the most difficulty adding to the left reported poor visual imagery but had good short term recall for numbers.

Product Complexity.

Once again product complexity emerged as a powerful factor, both in the CD analysis of variance and the CDIFF contrast correlations. Additions that yielded high complexity products were much more difficult than those that combined to form low complexity products. This held for both time, correctness, and the CDIFF score.

CDIFF contrast scores correlated with complex spatial tests (Paper Folding, Surface Development), Gf or G tests (Hidden Figures, Necessary Arithmetic Operations, Leter Series, Terman Concept Mastery, WAIS Block Design), and the WAIS Verbal tests. Thus, the effect once again appears to be associated with general ability. High ability students took more time and made more errors on high than on low product complexity additions.

The only significant CSUM correlation was with an "internal"

variable: Total Correct Draw (r=.44). This reflects the obvious conclusion that those who were more careful on high complexity additions were better able to draw them when requested to do so.

The N Contrast and Cell Totals.

The CDIFF N contrast reflects subjects' relative positions on a fast-accurate to slow-inaccurate scale. The only significant reference test correlations for this contrast were -.44 with Figures and -.58 with WAIS Picture Arrangement. Thus, fast-accurate construction was associated with high scores on these two tests.

The CSUM N contrast scores, on the other hand, correlated with a number of reference tests. The tests and their correlations with CSUM are shown in the first column of Table 3.14. High scores on these tests are associated with slow-accurate responding in this task. Given the preponderance of verbal tests in the list, it would appear that high CSUM scores were more associated with verbal ability (or Gc) than spatial ability (or Gf).

The N contrast scores for the independent analyses on the construction component time and SCORE3 were even more revealing. These contrasts represent average construction component time and average SCORE3 for each subject. As shown in Table 3.14, average construction time correlated with verbal tests, while average SCORE3 correlated with spatial tests.

Mean time and SCORE3 within each of the 24 cells of the construction component design were also correlated with the reference tests. Six of these component times correlated with WAIS Verbal Scale Score. All were for the more difficult cells, i.e., high stimulus complexity (four of six), high product complexity (all six), two piece additions (four of six), and left additions (four of six). Cell means for the easier items had small positive correlations with verbal tests and zero or slight negative correlations with spatial tests. However, only two Closure Speed test correlations were significant (see below).

For the SCORE3 cell totals, only four of the 24 cells

Table 3.14

Correlations between Construction CSUM, Component Time, and SCORE3 N Contrasts and Reference Tests (N=30)

Reference Test	CSUM	Time	SCORE3
Paper Folding	56		67
Form Board			55
Surface Development	48		63
Memory for Designs - Quality	58		72
Total Correct Draw Trials	76	51	73
Copying			44
Cards			47
Figures			55
Identical Pictures	41		45
Visual Number Span	59	52	41
Word Transformations	42		
Letter Series	50		45
Terman Concept Mastery	58	54	
WAIS Information	52	47	
Digit Span Forward	44	43	
Vocabulary	48	47	
Picture Arrangement			49
Object Assembly			42
Verbal Scale Score	52	45	
Performance Scale Score			57
Full Scale Score	52		52

Note. Decimals and correlations less than .41 omitted.

correlated significantly with WAIS Verbal Scale Score. However, these same cells also had high correlations with spatial tests, thus the relationship is with general ability. On the other hand, 23 cells correlated significantly with spatial tests. The remaining three cells did not correlate with anything in the battery.

Therefore, component latency and correctness were largely independent aspects of performance. Long solution times on difficult additions were associated with high verbal ability. Latency on easy additions was unrelated to either spatial or verbal ability. Correctness on both easy and difficult items was associated with spatial ability.

Closure Speed.

Twenty-one of the 24 correlations between construction component cell mean times and the Closure Speed reference factor were negative. However, only two were significant. The only design facet shared by the ten highest correlations (r=-.30 to -.60) was low product complexity. None of the SCORE3 cell total correlations with the Closure Speed factor were significant (in fact, only two were above .17). Thus, as argued elsewhere (Lohman, 1977b; 1979), Closure Speed is not a separate cognitive skill, but merely one aspect of spatial ability. Further, these results support the argument that the factor would be better represented by latency than the usual index of total correct within a given time limit.

The Rotation Component

Of the 216 items, 144 required rotation and thus had a rotation component time. Since construction always preceded rotation, rotation component times for the 96 items that also required construction were not experimentally independent of the construction operations. Rotation time on these items depended not only on the success of the rotation, but also on the success of the preceding addition.

Separate analyses were performed on the 48 items where

rotation was not preceded by construction and on the 96 items requiring both construction and rotation. Details of the latter analysis are presented in Lohman (1979b), and will be reviewed only briefly here.

No Construction Items.

Analyses of variance on rotation time and SCORE3 for the 48 no construction items are shown in Tables 3.15 and 3.16. Separate analyses were also performed on the sum and difference of normalized rotation time and correctness scores. However, since there were no discrepancies between the rotation time and SCORE3 analyses, the sum and difference scores provided no unique information. The difference score analysis coincided with those reported in Tables 3.15 and 3.16, while neither facets nor subject-facet interactions were significant in the sum score analysis.

Both rotation time and SCORE3 showed large main effects for rotation, no main effect for stimulus complexity, and a small, but significant rotation by stimulus complexity interaction. There were also large practice effects. Both rotation time and errors decreased linearly over the four blocks. Figure 3.6a shows a substantial decrease in rotation time over blocks, but only a slight increase in slopes. As is shown in Figure 3.6b, the decrease in errors was much larger for 180 degree rotations than for 90 degree rotations.

Although the rotation by stimulus complexity interaction was significant in both analyses, it is probably not reliable. The interaction for rotation time is plotted separately for each block in Figure 3.7. The interaction was different in each block, and significant only in the first two blocks.

Correlations

If nothing else, Figure 3.7 reveals that what was happening in Block 1 is not the same as what was happening in Blocks 3 and 4. This suggests that practice alters the relationships between rotation times and scores, and external variables.

Table 3.15

Analysis of Variance for Rotation Component Time

Source	df	MS	F	Source	df	MS	F[a]
Mean	1	3162.33	195.34*	N	29	16.18	-
Rotation	1	43.27	35.41*	NR	29	1.22	1.54
Stimulus Complexity	(2)			NS	(58)		
Linear	1	.05	.07	Linear	29	.83	1.05
Quadratic	1	1.44	3.43	Quadratic	29	.42	.53
RS	(2)			NRS	(58)		
Linear	1	1.99	2.11	Linear	29	.94	-
Quadratic	1	7.81	12.24*	Quadratic	29	.63	-

[a]MS(e) estimated by NRS interaction.

*p less than .01

Table 3.16

Analysis of Variance for Rotation SCORE3

Source	df	MS	F	Source	df	MS	F[a]
Mean	1	110.646	1121.19*	N	29	.099	-
Rotation	1	.176	11.15*	NR	29	.016	.81
Stimulus Complexity	(2)			NS	(58)		
Linear	1	.038	2.38	Linear	29	.016	.89
Quadratic	1	.002	.11	Quadratic	29	.019	.81
RS	(2)			NRS	(58)		
Linear	1	.022	1.01	Linear	29	.022	-
Quadratic	1	.207	11.96*	Quadratic	29	.017	-

[a]MS(e) estimated by the NRS interaction.

*p less than .01

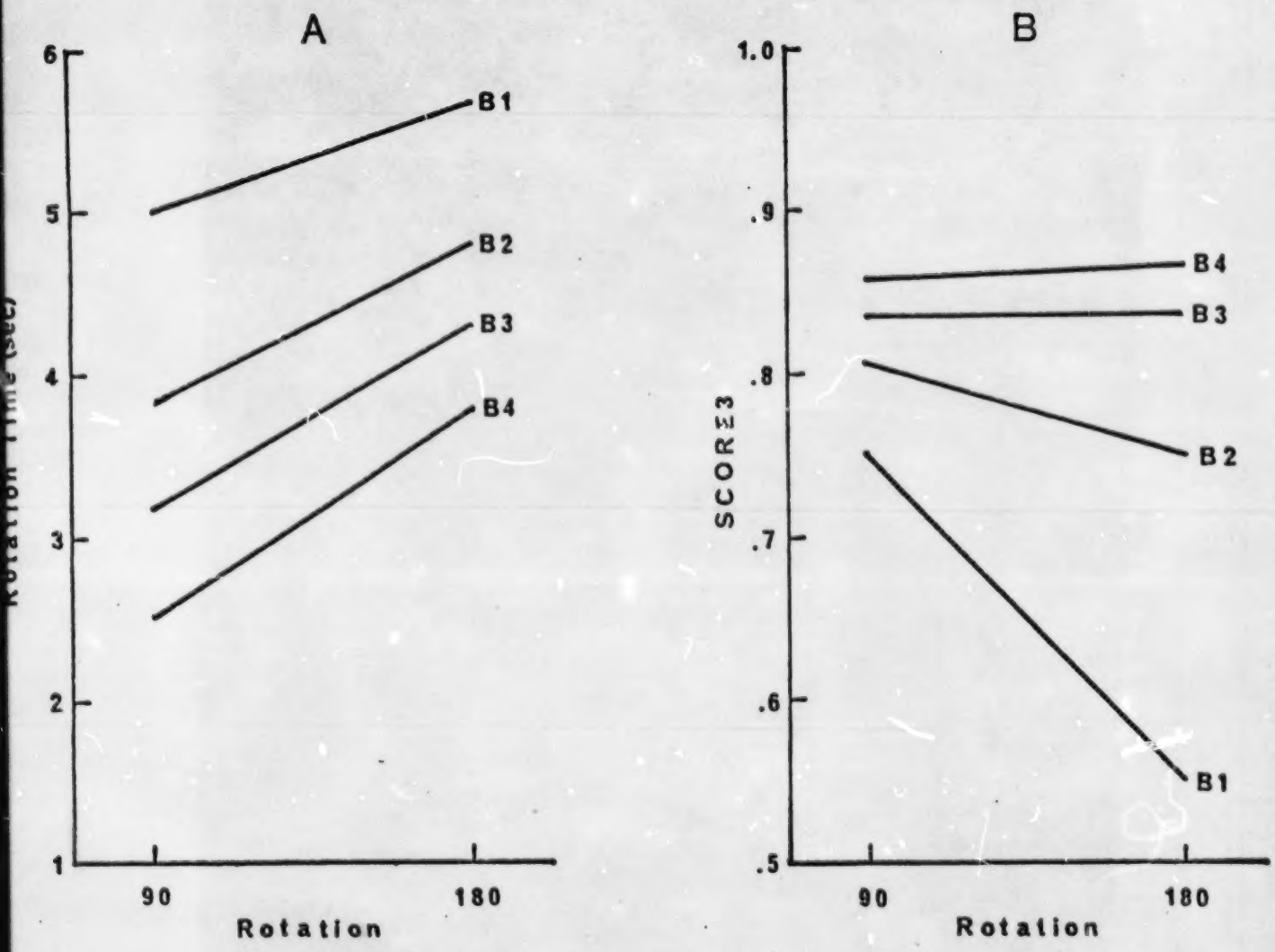

Figure 3.6. Plots of (a) rotation component time, and (b) SCORE3, by block for the 48 no construction items.

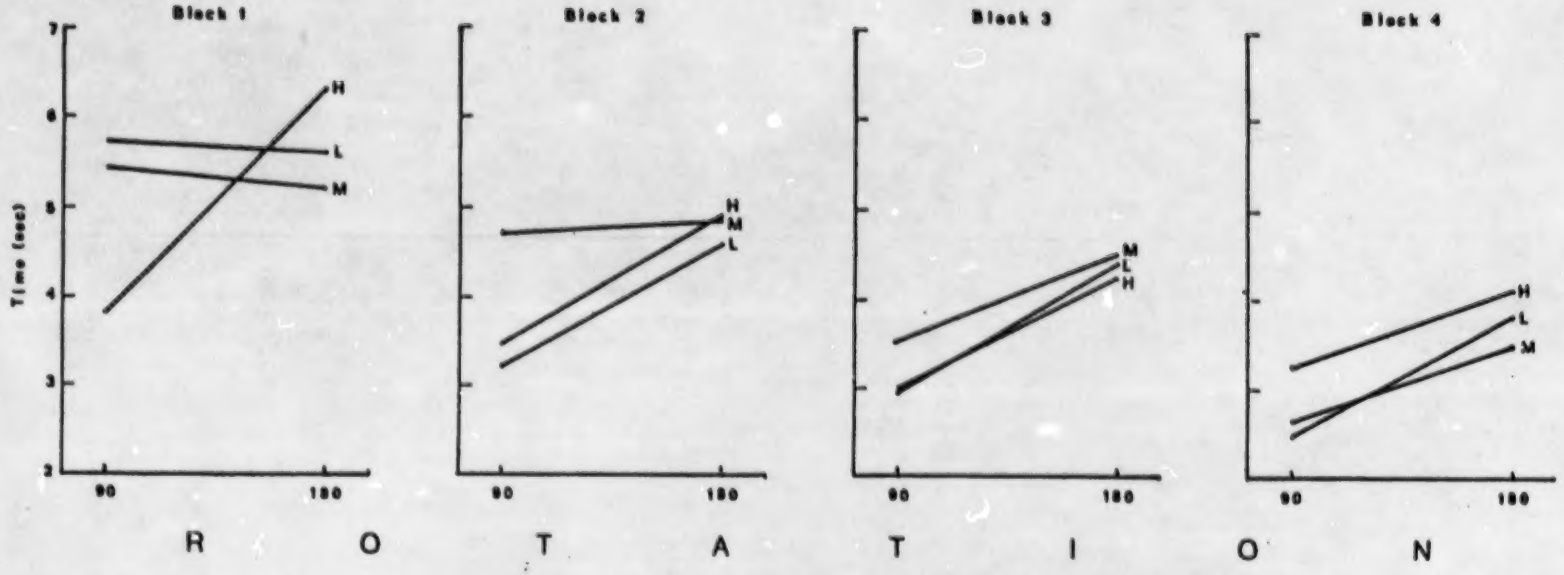

Figure 3.7. Plots of the rotation by stimulus complexity interaction, by block.

Correlations between mean SCORE and rotation time within each block, and selected reference variables are shown in Table 3.17. SCORE had its maximum correlations in Blocks 3 and 4, while rotation time correlations were highest in Blocks 1 and 2. While SCORE (and SCORE3) had many significant correlations with reference tests, mean rotation time correlated significantly only with number of errors on the Cards test.

When collapsed over blocks, number correct 180 degree rotations correlated slightly higher with reference tests than the number of correct 90 degree rotations. For example, correlations with Paper Folding were .46 and .53, and with Surface Development .44 and .54, for 90 and 180 degree rotations, respectively.

Once again, steep slopes were related to verbal ability. The rotation time contrast scores (180 time minus 90 time) correlated with WAIS Comprehension (.44), Digit Span (.50) and Verbal Scale Score (.46). For SCORE3, scores on the rotation contrast were positively correlated with two visual memory tests: Film Memory III (r=.53) and Visual Number Span (r=.51). Since Visual Number Span usually falls with the verbal tests or verbal strategy indices, and has frequently opposed the Marks Vividness of Visual Imagery Questionaire, the interpretation of these correlations is uncertain. Those who did not show a large decline in correctness between 90 and 180 degree rotations may have had good visual memory, and relied on it to rotate the entire stimulus, ignoring other cues that would make the task easier but more error prone.

The linear stimulus complexity scores for the rotation time contrast correlated .55 with Digit Span. Similarly, for SCORE3, the linear stimulus complexity contrast correlated -.52 with Digit Span and -.43 with Uses for Things. Thus, those who took longer and missed more high stimulus complexity rotations than low complexity rotations were probably attempting to encode and remember the stimulus or its distinctive features verbally.

Table 3.17

Correlations between Mean SCORE, Mean Rotation Component Time and Reference Tests, by Block (N=30)

	SCORE			
Reference Test	Block 1	Block 2	Block 3	Block 4
WAIS IQ	10	21	44*	26
Paper Folding	22	34	64*	44*
Surface Development	31	32	61*	35
Cards	02	26	39	44*
Figures	02	18	22	45*
Total Correct Draw Trials	28	43*	62*	41

	ROTATION TIME			
Reference Test	Block 1	Block 2	Block 3	Block 4
WAIS IQ	23	25	-13	-03
Cards Wrong	43*	55*	24	18
Identical Pictures	31	30	17	12
Total Correct Draw Trials	42*	44*	17	19

Note. Decimals omitted. r = .42 significant at .01

Rotation of Constructed Stimuli.

The major finding here was that it took longer to rotate a constructed stimulus. Figure 3.8 shows the plot of rotation time for each level of the construction facet. Product complexity did not affect rotation time, but did, of course, have large effects on error rate. Stimulus complexity affected both error rate and rotation time, especially for constructions involving high complexity stimuli. However, the previously noted limitations of these data make interpretations risky.

Match Component Times

Since accepting or rejecting the test stimulus was the last component operation in every item, match component times depend on the success of all previous operations. As such, they are the most difficult component times to interpret.

Separate analyses were performed on the 24 match lines preceded only by encoding and all 216 match times. Although analyses of variance were performed on match times and on the sum and difference of separately normalized match times and a correctness index, only descriptive statistics and correlational results are summarized here. Details of the analyses of variance are reported in Lohman (1979b).

All Match Times. Subjects were first divided into three groups on the basis of total SCORE's over all 216 items. Mean match time was then computed for each design facet separately for correct and incorrect responses for each group.

The major findings were:

1. Match times were shorter for correct responses than incorrect responses. Mean times were 1.43 and 1.76 sec for correct and incorrect responses, respectively. This is exactly the opposite of the results for the simple RT control. There, errors were consistently faster than correct responses.
2. Mean match time decreased linearly over the three groups, both for correct and incorrect responses.

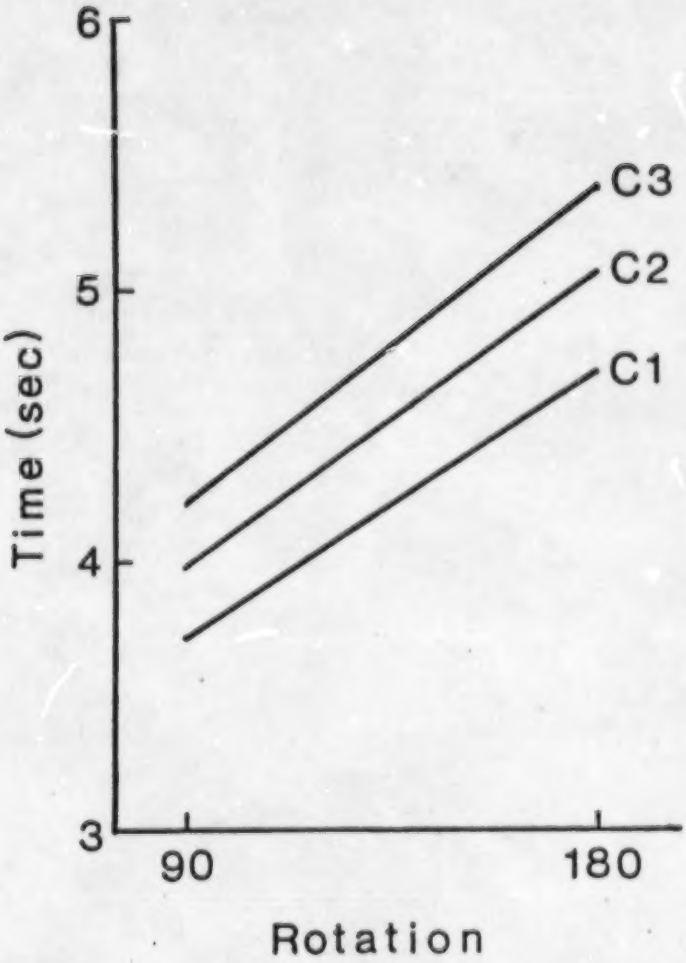

Figure 3.8. Plot of rotation component time by construction.

Further, incorrect match times were longer than correct times in all three groups. These data are shown in Figure 3.9.

3. There was a substantial interaction between correctness of the test probe and correctness of the response. These data are plotted in Figure 3.10a. "Yes" responses were faster on correct items, and "no" responses were faster on incorrect items. These differences were strongest in the low and medium groups.
4. Cell means for the rotation facet are plotted in Figure 3.10b. All three groups showed an increase in match time on rotation items for incorrect responses. However, only the low group showed a consistent increase for correct responses. In fact, both correct and incorrect match times increased linearly over levels of rotation in the low group. In the high group, on the other hand, correct match times were consistently rapid.
5. For construction, however, match time for all three groups were greater for construction than no construction items. As shown in Figure 3.10c, the most pronounced difference between groups was the magnitude of the increase between no construction and one piece additions.
6. Match times for correct responses increased linearly over both the stimulus complexity and type of addition facets. However, the increases were all less than one second.
7. Product complexity had large and consistent effects for all three groups. These means are plotted in Figure 3.10d.

Correlations.

Match time was the only component time that correlated significantly with spatial tests. Separate correlations with the reference tests were computed for mean match time over all 216

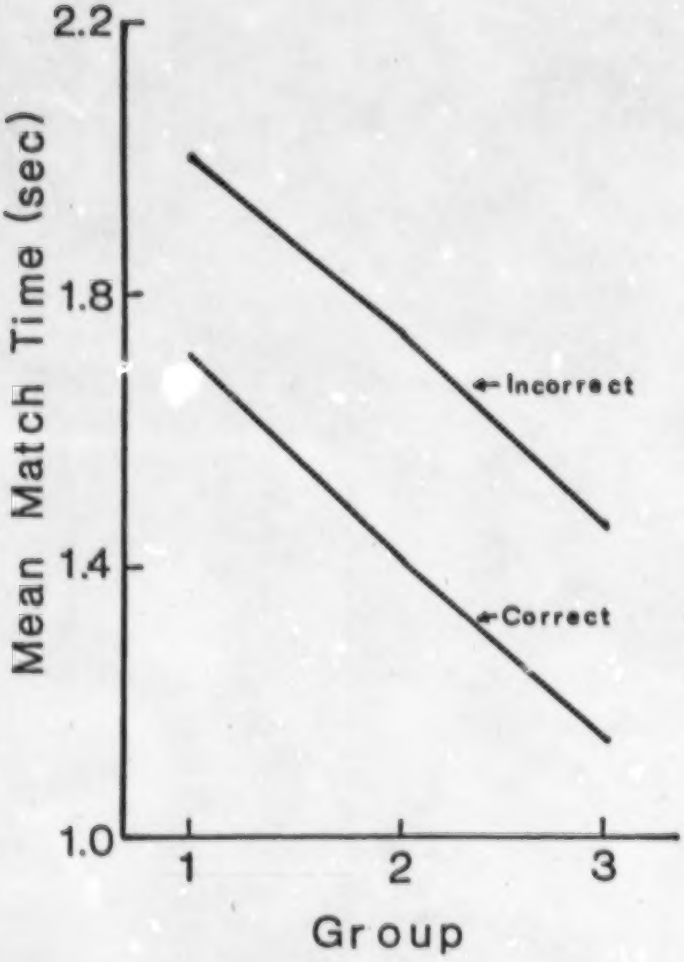

Figure 3.9. Mean match time for correct and incorrect items, by group.

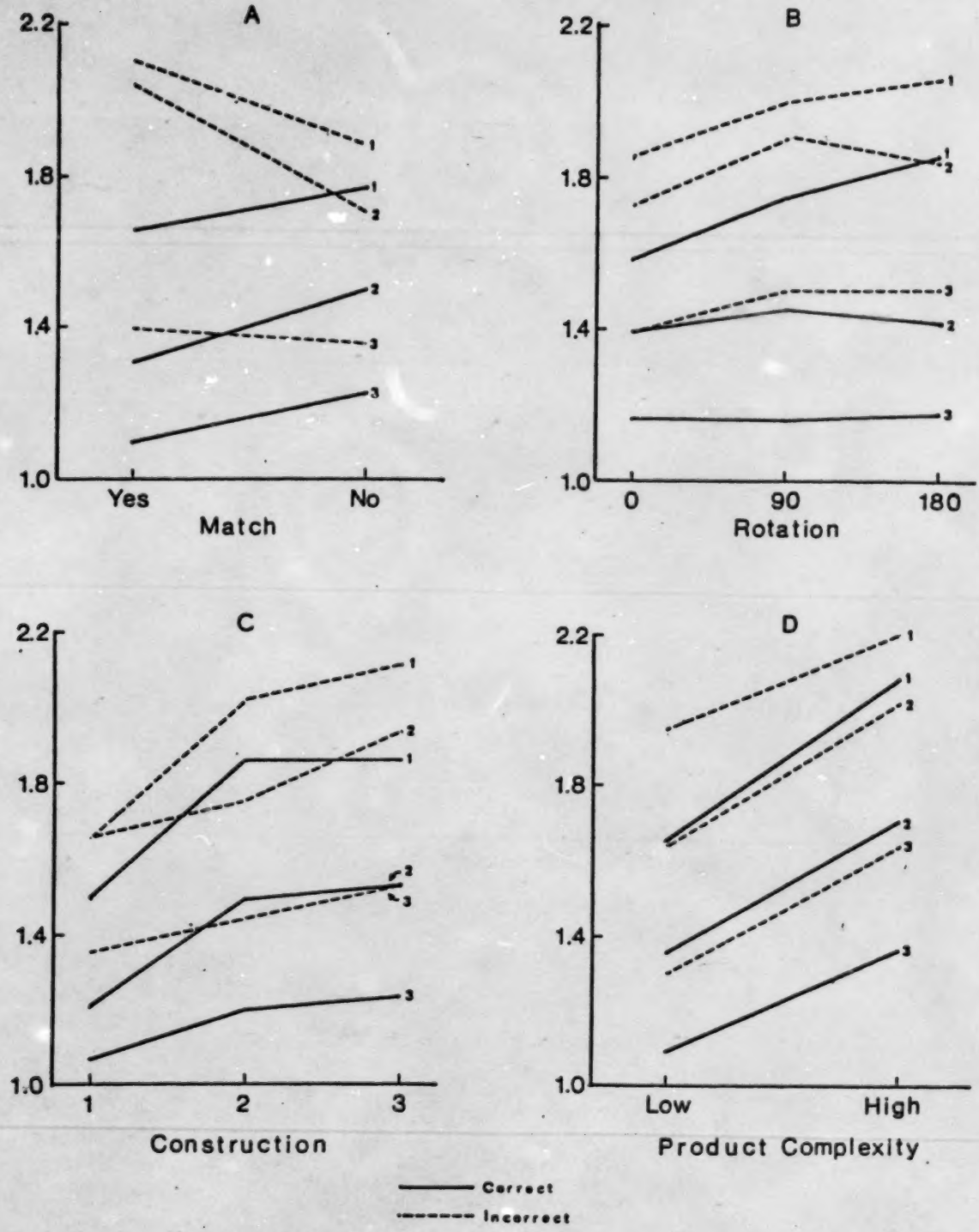

Figure 3.10. Match time for correct and incorrect items for (1) low, (2) medium, and (3) high performing groups, over (a) match, (b) rotation, (c) construction, and (d) product complexity.

items and mean match time over the 24 no construction, no rotation items. Four correlations were significant in one or both analyses. These correlations are shown in Table 3.18. The surprize here was that match time in the no construction, no rotation cells correlated with the reference spatial tests as well as overall match time. The mean error rate for the 24 no construction, no rotation items was only 4.4 percent, or one item. Thus, the correlation is not an artifact of the differential match times associated with correctness. Further, neither mean encoding time, mean total time (encoding time plus match time) nor total correct over these 24 items correlated significantly with any of the reference tests.

It was hypothesized that match times or total time on these simple no construction, no rotation items might be related to one or more of the perceptual speed tests, particularly Identical Pictures. However, mean match time on these items correlated only -.33, and total time -.02 with Identical Pictures.

Summary of Match Component Analyses.

Those who did not perform well on this task showed large increases in correct match times between no construction and construction items, and a linear increase in correct match time over the rotation facet. Those who performed well showed very little increase over the construction facet and no increase over rotation in correct match time. This could indicate that low ability subjects had not completed the construction or rotation operations before advancing to the match frame.

Two pieces of evidence argue against this interpretation. First, although construction times increased dramatically between one and two piece additions, match times did not. In fact, in the low performance group, mean correct match time was the same at levels two and three of the construction facet. On the other hand, mean correct match times increased linearly over the rotation facet within this low performance group. If these subjects were advancing to the next frame before completing the requested construction or rotation operation, then they did so

Table 3.18

Correlations between Mean Match Times and Reference Tests (N=30)

Reference Test	216 Items	24 Items
Form Board	-40	-44
Surface Development	-47	-39
Memory for Designs	-41	-48
Figures	-46	-43

Note. Decimals omitted.

differentially for construction and rotation.

Second, the highest correlation between mean match time over the 24 no construction, no rotation items and reference tests was -.48 with Memory for Designs. The crucial requirements of this test are to remember and reproduce the design in its correct proportions. Those who have difficulty retaining a given design in its correct proportions should have even greater difficulty keeping a constructed or transformed (i.e., rotated) figure in its correct proportions. The comparison and decision processes should take longer for these distorted images than for the relatively clear images where no construction or rotation was involved.

Smith (1964) has long argued that the essence of spatial ability is the ability to retain a stimulus in its correct proportions. The individual differences in match times suggest that this may indeed be an important aspect of spatial ability.

It must be emphasized, however, that the spatial skills tapped by the match times were largely independent of the level scores on this task. The correlation between mean match time over the 24 items and total SCORE was only -.24. Thus, while individual differences in match times were related to spatial ability, the basic hypothesis of speed-level independence remains unchallenged.

VI. Comparison of Overall and Component Analyses

This section compares and summarizes the major results of the total time and correctness, and component time and correctness analyses.

Total Scores

Individual differences in total correct (by SCORE3) were strongly correlated with spatial ability. Total time, on the other hand, was independent of spatial ability and weakly related to verbal ability.

Generally fast accurate performance on this task was correlated with only one reference test, WAIS Picture Arrangement. The tendency to opt for accuracy instead of speed

was related to general ability.

Encoding

Means Analysis.

Encoding component time decreased linearly over stimulus complexity and the four blocks of test trials. Individual differences in mean encoding time were highly correlated with the intercept parameters from the regressions of total and correct times on percent failing each item. Further, individual differences in mean encoding time were independent of the slope, predicted times at maximum difficulty, and all of the reference tests, except number of pencil marks made on the Form Board test.

Contrast Correlations.

Relatively large increases in encoding time over the stimulus complexity facet, especially between medium and high complexity, were associated with verbal ability and memory span. It was suggested that the reported strategies of attempting to remember a complex initial stimulus by viewing it as a combination of simpler forms, or attempting to associate the stimulus with a concrete image would result in relatively long encoding times for high complexity stimuli.

Change in encoding time over blocks was not related to the reference tests or overall performance on the experimental task.

Construction

Means Analysis.

Two piece additions took much longer than one piece additions, but did not result in more errors. On all other facets related to construction, error rate and total or component time increased together. The most difficult constructions were left additions of complex stimuli that combined to form high complexity products. Left additions were more difficult than right additions only when two pieces had to be added to the base. Left additions that resulted in high complexity products were more difficult than corresponding right additions. Difficulty

increased nonlinearly over stimulus complexity, with the difference between high and medium complexity larger than the corresponding difference between medium and low complexity stimuli. The largest changes in difficulty were captured by the product complexity facet, with high complexity products much more difficult to construct than low complexity products.

Contrast Correlations.

N Contrasts and Cell Totals. Mean construction component time correlated significantly with six verbal tests. Mean construction time on easy items was independent of the reference tests, while mean time on difficult items was highly correlated with verbal ability. Total SCORE3, on the other hand, was strongly related to performance on spatial tests, both for easy and difficult items.

Construction (C, K). In both the total time and construction component analyses, steep slopes were related to verbal ability. Further, the quadratic contrast in the total time analysis correlated positively with reported vividness of visual imagery and negatively with Hidden Figures. The linear SCORE3 construction contrast scores correlated positively with Form Board and Object Assembly, while the quadratic contrast scores had equally large, but negative correlations with these two tests. Thus, high scores on the Form Board and Object Assembly reference tests were related to a linear decline in performance over the three levels of the construction facet.

For the combined latency and correctness indices, larger increases in difficulty between one and two piece additions than between zero and one piece additions were, once again, related to Form Board and Object Assembly. However, magnitude of the simple difference between difficulty of one and two piece additions was related to verbal ability.

Shifting to slower, more accurate performance between levels one and three was associated with verbal ability, while relatively larger shifts between one and two piece additions than between zero and one piece additions were negatively related to

spatial ability. Thus, high spatial ability subjects showed the greatest change in speed-accuracy tradeoff between no addition and one addition items, rather than between one and two piece additions.

Stimulus Complexity. In both the total time and construction component analyses, the linear stimulus complexity contrast was related to verbal ability.

The quadratic contrast was strongly related to verbal, verbal reasoning, and perceptual speed tests in the total time analysis, but more related to general ability in the construction component time analysis.

Both linear and quadratic SCORE3 stimulus complexity constrasts were independent of the reference tests in the overall analysis, but weakly related to Copying and Figures in the analysis of the 144 construction items.

For the composite difficulty index, those who took much longer on constructions involving high complexity stimuli than on those using low complexity stimuli scored higher on a cluster of numerical or reasoning tests. A relatively larger increase in difficulty between medium and high complexity than between low and medium complexity was related to both spatial ability and memory span.

Strong shifts to slower, more accurate responding between low and high stimulus complexity additions was associated with high scores on word fluency tests.

Type of Addition. In both total time and construction component analyses, taking longer on left additions than right additions was related to good performance on Object Assembly and verbal tests. Further, those who missed many more left additions than right additions had high scores on the Visual Number Span test, while those whose performance was about the same on both reported strong visual imagery.

For the combined latency and correctness indices, more errors and longer latencies on left additions were even more strongly associated with high scores on Visual Number Span and low scores on the Marks Vividness of Visual Imagery Questionaire. Switching

to slower, more accurate responding on left additions was indicative of high scores on Form Board and Object Assembly.

Product Complexity. In both the total time and construction component time analyses, taking longer on high product complexity additions was strongly related to general ability. Time taken on low product complexity additions was also negatively correlated with Closure Speed. On the other hand, making relatively more errors on high complexity additions was weakly related to verbal ability.

The combined latency and correctness difficulty indices were also strongly related to general ability in both analyses. The magnitude of the speed-accuracy tradeoff, on the other hand, was uncorrelated with the reference tests and only weakly related to an "internal" variable, total number of correct draw trials.

Rotation

Means Analysis.

Total time increased nonlinearly while correctness decreased linearly over the rotation facet. The combined difficulty index, however, showed a perfectly linear increase. Further, there was a general shift to slower, more careful responding when rotation was required.

Both rotation time and errors decreased linearly over the four blocks. The decrease was largest for errors on 180 degree rotations.

Although the interaction between stimulus complexity and rotation was significant in both the component time and SCORE3 analyses, the interaction was different in each block and significant only in the first two blocks.

Rotation component time increased linearly over the construction facet. However, interpretation is difficult since construction always preceded rotation.

Contrast Correlations.

N Contrast. Mean rotation time correlated significantly with number of errors on the Cards test. When analyzed by block, the

correlation was significant only in the first block. Total number correct on rotation only items correlated significantly with spatial ability, particularly in blocks three and four. The highest correlations in block four were with Figures, Cards, and Paper Folding. Number of correct 180 degree rotations correlated slightly higher with spatial reference tests than number of correct 90 degree rotations.

Rotation. Steep slopes were associated with Cards Wrong in the total time analysis (R = 0, 90, 180), and verbal ability and Digit Span in the component time analysis (R = 90, 180). A linear decline in correctness was related to high scores on Figures and Cards in the overall analysis (R = 0, 90, 180), and high scores on two visual memory tests in the analysis of the 48 rotation only items (R = 90, 180).

Stimulus Complexity. A linear increase in rotation time and a linear decrease in correctness over stimulus complexity were related to high scores on memory span tests, particularly WAIS Digit Span Backwards. Further, those who made relatively more errors on rotations of high complexity stimuli performed well on Form Board, Object Assembly, and Memory for Designs.

Match

Means Analysis.

Match was the only nonsignificant main effect in the analysis of total time. This was the result of a substantial interaction between correctness of the test probe and correctness of the response. In particular, "yes" responses were faster on correct items and "no" responses faster on incorrect items. Those who performed well on the experimental task showed no increase in match time over the rotation facet and a small increase between construction and no construction items. Those who performed poorly showed a linear increase in match time over the rotation facet and a large increase in match time on construction items. Both correct and incorrect match times were consistently longer for high than for low product complexity stimuli.

Correlations.

For the match facet in the total time analysis, faster responses to positive test stimuli were related to high scores on the Form Board and Picture Completion tests. Faster responses to negative test stimuli were associated with more errors on the Cards test and higher scores on Visual Number Span.

Mean match component time was the only component time that correlated significantly with spatial ability. These correlations were the same for all 216 items and the 24 no construction, no rotation items. However, mean match time was not significantly correlated with level scores on the experimental task.

VII. General Discussion

The results of this experiment have several implications for research on aptitude processes.

The first important finding was that individual differences in speed of solving simple items were independent of level scores on the experimental task and reference tests. Thus, it is unlikely that a process understanding of spatial ability will be obtained by modeling individual differences in speed of solving simple spatial tasks.

The second important finding was that individual differences in the speed of solving complex tasks were consistently related to verbal reference tests. The implication here is that at least some subjects solved complex items using verbal-analytic strategies. This does not mean that all subjects used nonspatial strategies on complex items, or that individual differences in the speed of solving complex items are generally unrelated to spatial ability. The most likely possibility is that some subjects solve complex items on spatial tests using spatial strategies, others use verbal-analytic strategies, while most use both spatial and verbal processes. The high positive correlations between solution latency for complex items and verbal reference tests probably result from greater variance in

solution latency for those items solved by verbal-analytic strategies. This greater variance in latency in the nonspatial group would completely overshadow the much smaller variation in latency scores in the spatial group. Correlations for the total group would then tend to reflect the correlation in the verbal and mixed strategy groups.

The third important result was that level scores on this task were highly correlated with scores on a number of complex spatial tests in the reference battery. Thus, the findings are not limited to this particular task. This suggests that these reference spatial tests also can be solved in more than one way. Those who have studied retrospective and introspective reports of solution strategy on spatial tests have reached a similar conclusion (see Lohman, 1979a for a review). In short, the major challenge for any spatial test, whether simple or complex, is to keep it spatial.

There are two additional implications of the "nonspatial" nature of most spatial tests. First, verbal and spatial abilities may be more independent than the literature on mental tests suggests. Second, if some subjects solve items on a carefully constructed spatial test or experimental task using nonspatial strategies, then perhaps spatial ability is not that important in the real world (c.f. Cronbach, 1970), except, of course, when solution time is a crucial factor (e.g., piloting an airplane (Guilford & Lacey, 1947)). It is also possible that spatial ability tests would have higher validity coefficients if they could be made truly spatial. One simple way to do this would be to utilize both error and latency data, as in the SCORE3 index used in this study.

If some subjects solved some items on this experimental task using nonspatial strategies, then the obvious next question is who was using these nonspatial strategies, and on what items? Unfortunately, the experimental task was not designed to answer these questions. In the present experiment, the predicted patterns for latency and error data over the design facets are the same for both spatial and verbal-analytic solution

strategies. What is needed is an experimental task for which different solution strateagies would yield qualitatively different patterns of scores over specified design facets.

The most likely candidates for this type of experiment are the first and last component processes; namely encoding and discriminative response (or match). The encoding phase is important because the nature of the internal representation of the stimuli may determine, predispose, or at least limit the type of transformational processes that may operate on it. Thus, the subject who "burns in" a holistic representation is more likely to rotate the entire stimulus than one who encodes the same stimulus as "a rectangle with a small triangle cut out of the right side."

At the other end of the item, individual differences in component latencies for the discriminative response were the only latency measures that correlated with spatial ability. Further, scoring items with abnormally long discriminative response times incorrect was the major factor in making SCORE3 a "cleaner" measure of spatial ability than total number correct (SCORE). Cooper (1976) has also found important individual differences in discriminative response time on spatial tasks.

For both encoding and discriminative response component times, the crucial factor appears to be the nature of the internal representation the subject is either constructing or comparing with the test stimulus. Subjects who attempted to decompose complex stimuli into more familiar, labelable pieces took longer to encode, add, rotate, and compare the result with a test stimulus.

Another result that underscores the importance of the internal representation was that individual scores for the product complexity contrast were the best predictors of level scores on the experimental task. The interesting aspect of this facet is that it is a characteristic of the internal representation subjects are assumed to construct, and is independent of the complexity of the stimuli the subject actually sees. Again, however, subjects using both verbal-analytic and

spatial stratgies should show effects in the same direction over the facet.

This study has also pointed to some important methodological considerations for research on aptitude processes. First, it underscores the important differences between individual differences in speed and level. Inferences about the nature of general aptitude constructs based on experiments that model latency for relatively error free tasks are suspect.

Second, the method of computing sum and difference scores for the joint analyses of error and latency data, while superior to other methods that have been proposed (Pachella, 1974; Sternberg, 1977), is still inadequate. When a subject fails to solve an item, the reaction time for that item is ambiguous. Statistical manipulations cannot erase the psychological ambiguity of the data. A retreat to simple tasks and error free performance is not the answer, as this study has amply demonstrated. The important methodological realization is that models that account for performance on complex items are necessarily models for high ability subjects. Ability is confounded with item complexity because it is defined by the maximum difficulty of the items the subject can solve. Experimentally, then, a tailored testing approach should prove more useful than the typical procedure where all subjects attempt all items. The motivation of the low ability subjects should also improve since they would not be continually confronted with items that they cannot solve. Rectangular rather than representative sampling of subjects could then be used to insure the same number of high, medium, and low ability subjects in the design. Low ability subjects would provide unambiguous latency data for easy items, medium ability subjects for easy and medium difficulty items, and high ability subjects for easy, medium, and difficult items. The analysis could then contrast the way low, medium, and high ability subjects solve easy items; the way medium and high ability subjects solve medium difficulty items and the way high ability subjects solve difficult items. Contrasts within an ability group could also be made, such as between the models that best

fit the performance of high ability subjects on easy, medium, and difficult items.

Another implication for future research stems from the conclusion that subjects solve different items in different ways. The analysis must go beyond fitting one model for each subject. It must focus on the way a particular subject solves a particular item type. Level of practice, item difficulty, and susceptibility of the item to alternative solution strategies are three possible blocking factors for such a design.

CHAPTER IV

SUMMARY

Previous research on the nature of aptitude processes has ignored the important differences between speed and level of performance. However, investigations of the relationship between individual differences in speed and level suggest that speed of solving simple spatial tasks is largely independent of level attained on more complex spatial tasks.

The present experiment was designed to investigate the relationships between speed, level, and complexity in individual differences in spatial ability. Thirty male high school and college students were selected to represent a wide range of individual differences in verbal ability and spatial ability. Subjects were required to match, rotate, combine, and combine and then rotate three to eight point polygons. Three levels of construction (one, two, or three stimuli), three levels of stimulus complexity (low, medium, or high), three levels of rotation (0, 90, or 180 degrees), and two types of discriminative response (correct or incorrect) were fully crossed in the design. Additionally, type of addition (left or right) and complexity of the to-be-constructed stimuli were nested within levels of the construction facet but crossed with each other and all other facets. Since each subject was administered all 216 items, subjects were the seventh fully crossed factor.

Thirty trials on which the subject was asked to draw the final stimulus rather than merely to recognize it were randomly interspersed with the test trials. A simple reaction time control condition preceded the experiment.

Correctness, confidence, and from two to four component times were obtained on each item. Total time was determined by summing the component times. These data were analyzed using regression, analysis of variance, and correlational techniques.

In the regression analyses, latency measures were regressed on a group difficulty scale. Correlations between intercepts, slopes, predicted times at maximum difficulty, and level scores and reference constructs supported the basic hypothesis of speed-

level independence. In particular, individual differences in speed of solving simple items were independent of both level on the task and reference abilities. Speed of solving complex items was strongly correlated with level on the task, but more highly correlated with verbal than spatial reference constructs.

In the means analysis, repeated measures analyses of variance were performed on total time and correctness, component times and correctness, and combined indices representing the sum and difference of independently standardized latency and correctness scores. These combined analyses were performed on both total time and correctness as well as each component time and its correctness index.

Relationships between individual differences in total scores, cell totals, and the pattern of cell means over design facets were explored in the correlational analyses. In general, correlations between correctness indices and reference tests supported the construct validity of the task and its facets, while correlations between latency indices and reference tests suggested more about the processes or strategies subjects were using to solve the items.

The major results of the means analyses and correlational analyses are summarized below.

Total Scores

Individual differences in total correct (by SCORE3) were strongly correlated with spatial ability. Total time, on the other hand, was independent of spatial ability and weakly related to verbal ability.

Generally fast accurate performance on this task was correlated with only one reference test, WAIS Picture Arrangement. The tendency to opt for accuracy instead of speed was related to general ability.

Encoding

Means Analysis.

Encoding component time decreased linearly over stimulus

complexity and the four blocks of test trials. Individual differences in mean encoding time were highly correlated with the intercept parameters from the regressions of total and correct times on percent failing each item. Further, individual differences in mean encoding time were independent of the slope, predicted times at maximum difficulty, and all of the reference tests, except number of pencil marks made on the Form Board test.

Contrast Correlations.

Relatively large increases in encoding time over the stimulus complexity facet, especially between medium and high complexity, were associated with verbal ability and memory span. It was suggested that the reported strategies of attempting to remember a complex initial stimulus by viewing it as a combination of simpler forms, or attempting to associate the stimulus with a concrete image would result in relatively long encoding times for high complexity stimuli.

Change in encoding time over blocks was not related to the reference tests or overall performance on the experimental task.

Construction

Means Analysis.

Two piece additions took much longer than one piece additions, but did not result in more errors. On all other facets related to construction, error rate and total or component time increased together. The most difficult constructions were left additions of complex stimuli that combined to form high complexity products. Left additions were more difficult than right additions only when two pieces had to be added to the base. Left additions that resulted in high complexity products were more difficult than corresponding right additions. Difficulty increased nonlinearly over stimulus complexity, with the difference between high and medium complexity larger than the corresponding difference between medium and low complexity stimuli. The largest changes in difficulty were captured by the product complexity facet, with high complexity products much more

difficult to construct than low complexity products.

Contrast Correlations.

N Contrasts and Cell Totals. Mean construction component time correlated significantly with six verbal tests. Mean construction time on easy items was independent of the reference tests, while mean time on difficult items was highly correlated with verbal ability. Total SCORE3, on the other hand, was strongly related to performance on spatial tests, both for easy and difficult items.

Construction (C, K). In both the total time and construction component analyses, steep slopes were related to verbal ability. Further, the quadratic contrast in the total time analysis correlated positively with reported vividness of visual imagery and negatively with Hidden Figures. The linear SCORE3 construction contrast scores correlated positively with Form Board and Object Assembly, while the quadratic contrast scores had equally large, but negative correlations with these two tests. Thus, high scores on the Form Board and Object Assembly reference tests were related to a linear decline in performance over the three levels of the construction facet.

For the combined latency and correctness indices, larger increases in difficulty between one and two piece additions than between zero and one piece additions were, once again, related to Form Board and Object Assembly. However, magnitude of the simple difference between difficulty of one and two piece additions was related to verbal ability.

Shifting to slower, more accurate performance between levels one and three was associated with verbal ability, while relatively larger shifts between one and two piece additions than between zero and one piece additions were negatively related to spatial ability. Thus, high spatial ability subjects showed the greatest change in speed-accuracy tradeoff between no addition and one addition items, rather than between one and two piece additions.

Stimulus Complexity. In both the total time and construction

component analyses, the linear stimulus complexity contrast was related to verbal ability.

The quadratic contrast was strongly related to verbal, verbal reasoning, and perceptual speed tests in the total time analysis, but more related to general ability in the construction component time analysis.

Both linear and quadratic SCORE3 stimulus complexity constrasts were independent of the reference tests in the overall analysis, but weakly related to Copying and Figures in the analysis of the 144 construction items.

For the composite difficulty index, those who took much longer on constructions involving high complexity stimuli than on those using low complexity stimuli scored higher on a cluster of numerical or reasoning tests. A relatively larger increase in difficulty between medium and high complexity than between low and medium complexity was related to both spatial ability and memory span.

Strong shifts to slower, more accurate responding between low and high stimulus complexity additions was associated with high scores on word fluency tests.

Type of Addition. In both total time and construction component analyses, taking longer on left additions than right additions was related to good performance on Object Assembly and verbal tests. Further, those who missed many more left additions than right additions had high scores on the Visual Number Span test, while those whose performance was about the same on both reported strong visual imagery.

For the combined latency and correctness indices, more errors and longer latencies on left additions were even more strongly associated with high scores on Visual Number Span and low scores on the Marks Vividness of Visual Imagery Questionaire. Switching to slower, more accurate responding on left additions was indicative of high scores on Form Board and Object Assembly.

Product Complexity. In both the total time and construction component time analyses, taking longer on high product complexity additions was strongly related to general ability. Time taken on

low product complexity additions was also negatively correlated with Closure Speed. On the other hand, making relatively more errors on high complexity additions was weakly related to verbal ability.

The combined latency and correctness difficulty indices were also strongly related to general ability in both analyses. The magnitude of the speed-accuracy tradeoff, on the other hand, was uncorrelated with the reference tests and only weakly related to an "internal" variable, total number of correct draw trials.

Rotation

Means Analysis.

Total time increased nonlinearly while correctness decreased linearly over the rotation facet. The combined difficulty index, however, showed a perfectly linear increase. Further, there was a general shift to slower, more careful responding when rotation was required.

Both rotation time and errors decreased linearly over the four blocks. The decrease was largest for errors on 180 degree rotations.

Although the interaction between stimulus complexity and rotation was significant in both the component time and SCORE3 analyses, the interaction was different in each block and significant only in the first two blocks.

Rotation component time increased linearly over the construction facet. However, interpretation was difficult since construction always preceded rotation.

Contrast Correlations.

N Contrast. Mean rotation time correlated significantly with number of errors on the Cards test. When analyzed by block, the correlation was significant only in the first block. Total number correct on rotation only items correlated significantly with spatial ability, particularly in blocks three and four. The highest correlations in block four were with Figures, Cards, and Paper Folding. Number of correct 180 degree rotations correlated

slightly higher with spatial reference tests than number of correct 90 degree rotations.

Rotation. Steep slopes were associated with Cards Wrong in the total time analysis (R = 0, 90, 180), and verbal ability and Digit Span in the component time analysis (R = 90, 180). A linear decline in correctness was related to high scores on Figures and Cards in the overall analysis (R = 0, 90, 180), and high scores on two visual memory tests in the analysis of the 48 rotation only items (R = 90, 180).

Stimulus Complexity. A linear increase in rotation time and a linear decrease in correctness over stimulus complexity were related to high scores on memory span tests, particularly WAIS Digit Span Backwards. Further, those who made relatively more errors on rotations of high complexity stimuli performed well on Form Board, Object Assembly, and Memory for Designs.

Match

Means Analysis.

Match was the only nonsignificant main effect in the analysis of total time. This was the result of a substantial interaction between correctness of the test probe and correctness of the response. In particular, "yes" responses were faster on correct items and "no" responses faster on incorrect items. Those who performed well on the experimental task showed no increase in match time over the rotation facet and a small increase between construction and no construction items. Those who performed poorly showed a linear increase in match time over the rotation facet and a large increase in match time on construction items. Both correct and incorrect match times were consistently longer for high than for low product complexity stimuli.

Correlations.

For the match facet in the total time analysis, faster responses to positive test stimuli were related to high scores on the Form Board and Picture Completion tests. Faster responses to negative test stimuli were associated with more errors on the

Cards test and higher scores on Visual Number Span.

Mean match component time was the only component time that correlated significantly with spatial ability. These correlations were the same for all 216 items and the 24 no construction, no rotation items. However, mean match time was not significantly correlated with level scores on the experimental task.

CHAPTER V
CONCLUSIONS

1. The most important result of this experiment was the finding that individual differences in speed of doing simple spatial tasks are unrelated to verbal ability, spatial ability, or level scores on the experimental task. Speed of solving complex items was highly correlated with two task level scores and verbal reference tests.

2. In general, individual differences in the patterns of correctness cell means over facets supported the construct validity of the facet, while individual differences in the patterns of latency cell means over facets suggested more about the processes or strategies subjects were using to solve the items. Correlations between individual contrast scores and reference tests provided a powerful method for examining these relationships.

3. Large interactions between facets implied that a simple additive mathematical model or process model that postulated independent stages would not describe these data well. Complexity or difficulty increased nonlinearly.

4. Since time and correctness related differently to external variables, a complete analysis must examine both. A method designed to analyze their joint effects was proposed. Time and correctness were first independently standardized to zero mean and unit variance. The sum and difference of these two scores were then computed. The difference score represented a composite index of difficulty, while the sum score reflected the unique contribution of each variable, or variations in speed-accuracy tradeoff. Although new relationships sometimes emerged when contrast scores for these variables were correlated with reference constructs, it was concluded that correlational analyses were best performed on the original latency and correctness variables. The composite variables were, however,

quite appropriate for the means analysis.

5. The relationships between overall performance, performance on a particular item type, the patterns of performance over facets, and reference tests provided insights into the nature of the task and the nature of the reference constructs. Total number correct was strongly correlated with a cluster of complex spatial tests called Visualization in the factor analytic literature. Performance on the construction facet was consistently related to Form Board and Object Assembly, suggesting that mental construction may indeed be an important spatial skill. Similarly, performance over the rotation facet was related to performance on two tests (Cards and Figures) that usually define the factor called Spatial Relations. The Closure Speed factor was related to time spent on constructions that yielded a low complexity product. Other clusters of tests representing such diverse factors as Reasoning (or Number), Word Fluency, Visual Memory, and Memory Span were also related to particular patterns of performance on the task.

6. Several important aspects of the spatial construction task were identified. In particular, mental construction was differentially affected by the number of stimulus elements to be combined, the complexity of these stimuli, the location (or type) of the addition, and, most significantly, the complexity of the to-be-constructed stimulus. Further, large increases in latency over construction, stimulus complexity, and type of addition were related to verbal ability and verbal solution strategies. Rapid construction on low product complexity items was related to Closure Speed, while good overall performance was related to Form Board and Object Assembly. Finally, differential correlations with Hidden Figures and Marks Vividness of Visual Imagery Questionaire suggested that holistic and analytic spatial construction strategies must be distinguished.

7. Analysis of the rotation items confirmed Cooper and

Podgorny's (1976) finding that rate of rotation was independent of stimulus complexity. Examination of individual differences added that steep reaction time slopes for rotation were related to verbal ability, short term verbal memory, and verbal strategies. Further, large increases in rotation time for complex stimuli were also associated with verbal memory span tests.

8. Some subjects attempted to solve some items on this task using nonspatial strategies. However, since total SCORE's were highly correlated with performance on other complex spatial tests, these tests are also suspect. Those who have studied retrospective reports of solution strategies on spatial tests have reached the same conclusion (see Lohman, 1979a, for a review). The problem is particularly acute on complex spatial tasks. Some subjects solve easy items using spatial strategies, but find that these strategies break down as item difficulty increases. They then resort to verbal-analytic strategies. Strategies also change with practice. Attempting to eliminate individual differences in solution strategy by keeping the task simple will not do, as simple tests are also solved in different ways by different subjects. Further, as this study has shown, individual differences in speed of solving simple spatial tasks are largely independent of more generalizable level scores on complex tasks of the same type. Future research on spatial ability must attend to these important intra-subject and inter-subject differences in solution strategy. One way to do this is through experiments that are designed to yield qualitatively different patterns of scores for different solution strategies. Proportional subject sampling, selective item administration, and modeling performance at the level of item types within individuals are also recommended as part of the proposed research strategy.

References

Arnoult, M. D. Prediction of perceptual responses from structural characteristics of the stimulus. Perceptual and Motor Skills, 1960, 11, 261-268.

Attneave, F. Physical determinants of the judged complexity of shapes. Journal of Experimental Psychology, 1957, 53, 221-227.

Attneave, F., & Arnoult, M. D. The quantative study of shape and pattern perception. Psychological Bulletin, 1956, 53, 452-471.

Barratt, E. S. An analysis of verbal reports of solving spatial problems as an aid in defining spatial factors. Journal of Psychology, 1953, 36, 17-25.

Calfee, R. C. Sources of dependency in cognitive processes. In D. Klahr (Ed.), Cognition and instruction. Hillsdale, N. J.: Lawrence Erlbaum Assoc., 1976.

Cattell, R. B. Theory of fluid and crystallized intelligence: A critical experiment. Journal of Educational Psychology, 1963, 54, 1-22.

Clark, H., & Chase, W. On the process of comparing sentences against pictures. Cognitive Psychology, 1972, 3, 472-517.

Cooper, L. A. Mental transformation of random two-dimensional shapes. Cognitive Psychology, 1975, 7, 20-43.

Cooper, L. A. Individual differences in visual comparison processes. Perception & Psychophysics, 1976, 19, 433-444.

Cooper, L. A., & Podgorny, P. Mental transformations and visual comparison processes: Effects of complexity and similarity. Journal of Experimental Psychology: Human Perception and Performance, 1976, 2, 503-514.

Cooper, L. A., & Shepard, R. N. Chronometric studies of the rotation of mental images. In W. G. Chase (Ed.), Visual information processing. New York: Academic Press, 1973. (a)

Cooper, L. A., & Shepard, R. N. The time to prepare for a rotated stimulus. Memory & Cognition, 1973, 1, 246-250. (b)

Cooper, L. A., & Shepard, R. N. Transformations on representations of objects in space. Technical Report No. 59. Center for Information Processing, University of California, San Diego, 1976.

Cronbach, L. J. Essentials of psychological testing (3rd Edition). New York: Harper and Row, 1970.

Cronbach, L. J., & Furby, L. How do we measure "change"- or should we? Psychological Bulletin, 1970, 74, 68-80.

Cronbach, L. J., & Snow, R. E. Aptitudes and instructional methods: A handbook for research on interactions. New York: Irvington, 1977.

French, J. W. The description of aptitude and achievement tests in terms of rotated factors. Psychometric Monograph No. 5, 1951.

French, J. W. The relationship of problem-solving styles to the factor composition of tests. Educational and Psychological Measurement, 1965, 15, 9-28.

French, J. W., Ekstrom, R. B., & Price, L. A. Kit of reference tests for cognitive factors. Princeton, N. J.: Educational Testing Service, 1963.

Furneaux, W. D. Intellectual abilities and problem-solving behavior. In H. J. Eysenck (Ed.), Handbook of abnormal psychology. New York: Basic Books, 1961.

Garner, W. R. The processing of information and structure. Potomac, Md.: Erlbaum, 1974.

Garner, W. R., & Clement, D. E. Goodness of pattern and pattern uncertainty. Journal of Verbal Learning and Verbal Behavior, 1963, 2, 446-452.

Getzels, J. W., & Jackson, P. W. Creativity and intelligence. New York: Wiley, 1962.

Glushko, R. J., & Cooper, L. A. Constructing internal representations from structural descriptions. Paper in preparation, 1976.

Graham, F. K., & Kendall, B. S. Memory for designs test. Department of Neuropsychiatry, Washington University School of Medicine, 1948.

Guilford, J. P. The nature of human intelligence. New York: McGraw-Hill, 1967.

Guilford, J. P., & Hoepfner, R. The analysis of intelligence. New York: McGraw-Hill, 1971.

Guilford, J. P., & Lacey, J. I. (Eds.) Printed classification tests. AAF aviation psychology research program reports, No. 5, Washington, D. C.: Government Printing Office, 1947.

Harshman, R. Harshman Figures. Los Angeles: Department of Linguistics, University of California at Los Angeles, 1974.

Horn, J. L. Human abilities: A review of research and theory in the early 1970's. Annual Review of Psychology, 1976, 27, 437-485.

Hunt, E. B., Frost, N., & Lunneborg, C. E. Individual differences in cognition: A new approach to intelligence. In G. Bower (Ed.), Advances in learning and memory (Vol. 7). New York: Academic Press, 1973.

Kotovsky, K., & Simon, H. A. Empirical tests of a theory of human acquisition of concepts for sequential events. Cognitive Psychology, 1973, 4, 399-424.

Koussy, A. A. H. El. The visual perception of space. British Journal of Psychology, Monograph Supplement No. 20, 1935.

Lohman, D. F. Eye movement differences reflecting aptitude processes. Paper presented at the meeting of the American Psychological Association, San Francisco, August 1977. (a)

Lohman, D. F. The relationship between hypnotizability and speed of closure. Technical Report No. 6, Aptitude Research Project, School of Education, Stanford University, 1977. (b)

Lohman, D. F. Spatial ability: A review and reanalysis of the correlational literature. Technical Report No. 8, Aptitude Research Project, School of Education, Stanford University, 1979. (a)

Lohman, D. F. Individual differences in speed and level in spatial ability. Unpublished doctoral dissertation, Stanford University, 1979. (b)

Lord, F. M. Elementary models for measuring change. In C. W. Harris (Ed.), Problems in measuring change. Madison: University of Wisconsin Press, 1963.

Marks, D. F. Visual imagery differences in the recall of pictures. British Journal of Psychology, 1973, 64, (1), 17-24.

Metzler, J., & Shepard, R. N. Transformational Studies of the internal representation of three-dimensional objects. In R. Solso (Ed.), Theories in cognitive psychology: The Loyola Symposium. Potomac, Md.: Lawrence Erlbaum, 1974, pp 147-201.

Meyers, C. T. Some observations of problem solving in spatial relations tests. RB-58-16, Princeton, N. J.: Educational Testing Service, 1958.

Nisbett, R. E., & Wilson, T. D. Telling more than we can know: Verbal reports on mental processes. Psychological Review, 1977, 84, 231-259.

Pachella, R. G. The interpretation of reaction time in information processing research. In B. Kantowitz (Ed.), Human information processing: Tutorials in performance and cognition. Hillsdale, W. J.: Lawrence Erlbaum Assoc., 1974.

Palmer, S. E. Structural aspects of perceptual organization. Unpublished doctoral dissertation, University of California, San Diego, 1974.

Podgorny, P. Mental Rotation and the third dimension. Unpublished senior honors thesis, Stanford University, 1975.

Raven, J. C. Advanced progressive matrices, New York: Psychological Corporation, 1962.

Seibert, W. F., & Snow, R. E. Studies in cine-psychometry I: Preliminary factor analysis of visual cognition and memory (Final Report). Lafayette, Purdue University, July 1965.

Shepard, R. N. Form, formation, and transformation of internal representations. In R. Solso (Ed.), Information processing and cognition: The Loyola Symposium. Hillsdale, N. J.: Lawrence Erlbaum, 1975.

Shepard, R. N., & Feng, C. A chronometric study of mental paper folding. Cognitive Psychology, 1972, 3, 228-243.

Shepard, R. N., & Metzler, J. Mental rotation of three-dimensional objects. Science, 1971, 171, 701-703.

Simon, H. A., & Kotovsky, D. Human acquisition of concepts for sequential patterns. Psychological Review, 1963, 70, 534-546.

Smith, I. M. Spatial ability. San Diego, Calif.: Knapp, 1964.

Snow, R. E. Theory and method for research on aptitude processes: A prospectus. Technical Report No. 2, Aptitude Research Project, School of Education, Stanford University, 1976.

Snow, R. E. Individual differences, instructional theory, and instructional design. Technical report No. 4, Aptitude Research Project, School of Education, Stanford University, 1977.

Snow, R. E., Marshalek, B., & Lohman, D. F. Correlation of selected cognitive abilities and cognitive processing parameters: An exploratory study. Technical Report No. 3, Aptitude Research Project, School of Education, Stanford University, 1976.

Stenson, H. H. The physical factor structure of random forms and their judged complexity. Perception & Psychophysics, 1966, 1, 303-310.

Sternberg, R. J. Intelligence, information processing and analogical reasoning: The componential analysis of human abilities. Hillsdale, N. J.: Lawrence Erlbaum, 1977.

Street, R. F. A geslalt completion test. Teachers college, Columbia University Thesis, No. 4, 1931.

Terman, L. Concept Mastery Test (Form T). New York: Psychological Corporation, 1950.

Thorndike, E. L. The measurement of intelligence. Bureau of Publications, Teachers College, Columbia University,1926.

Thurstone, L. L. Ability, motivation, and speed. Psychometrica, 1937, 2, 249-254.

Thurstone, L. L. Primary mental abilities. Psychometric Monographs, No. 1, 1938.

Wechsler, D. Wechsler adult intelligence scale. New York: Psychological Corporation, 1955.

Witkin, H. A., Oltman, P. K., Raskin, E., & Karp, S. A. Embedded figures test. Palo Alto, Ca.: Consulting Psychologists Press, 1971.

Zimmerman, W. S. A revised orthogonal rotational solution for Thurstone's Primary Mental Abilities test battery. Psychometrica, 1953 18, 77-93.

DISTRIBUTION LIST

Navy

1 Dr. Ed Aiken
Navy Personnel R&D Center
San Diego, CA 92152

1 Dr. Jack R. Borsting
Provost & Academic Dean
U.S. Naval Postgraduate School
Monterey, CA 93940

1 DR. JOHN F. BROCK
NAVY PERSONNEL R& D CENTER
SAN DIEGO, CA 92152

1 MR. MAURICE CALLAHAN
Pers 23a
Bureau of Naval Personnel
Washington, DC 20370

1 CAPT. H. J. CONNERY
NAVY MEDICAL R&D COMMAND
NNMC
BETHESDA, MD 20014

1 Dr. Richard Elster
Department of Administrative Sciences
Naval Postgraduate School
Monterey, CA 93940

1 DR. PAT FEDERICO
NAVY PERSONNEL R&D CENTER
SAN DIEGO, CA 92152

1 CDR John Ferguson, MSC, USN
Naval Medical R&D Command (Code 44)
National Naval Medical Center
Bethesda, MD 20014

1 Dr. Paul Foley
Navy Personnel R&D Center
San Diego, CA 92152

1 Dr. John Ford
Navy Personnel R&D Center
San Diego, CA 92152

1 CAPT. D.M. GRAGG, MC, USN
HEAD, SECTION ON MEDICAL EDUCATION
UNIFORMED SERVICES UNIV. OF THE
HEALTH SCIENCES
6917 ARLINGTON ROAD
BETHESDA, MD 20014

1 MR. GEORGE N. GRAINE
NAVAL SEA SYSTEMS COMMAND
SEA 05L1C2
WASHINGTON, DC 20362

1 CDR Robert S. Kennedy
Naval Aerospace Medical and
Research Lab
Box 29407
New Orleans, LA 70189

1 Dr. Norman J. Kerr
Chief of Naval Technical Training
Naval Air Station Memphis (75)
Millington, TN 38054

1 Dr. Leonard Kroeker
Navy Personnel R&D Center
San Diego, CA 92152

1 CHAIRMAN, LEADERSHIP & LAW DEPT.
DIV. OF PROFESSIONAL DEVELOPMMENT
U.S. NAVAL ACADEMY
ANNAPOLIS, MD 21402

1 Dr. William L. Maloy
Principal Civilian Advisor for
Education and Training
Naval Training Command, Code 00A
Pensacola, FL 32508

1 CAPT Richard L. Martin
USS Francis Marion (LPA-Z49)
FPO New York, NY 09501

1 Dr. James McBride
Code 301
Navy Personnel R&D Center
San Diego, CA 92152

1 DR. WILLIAM MONTAGUE
LRDC
UNIVERSITY OF PITTSBURGH
3939 O'HARA STREET
PITTSBURGH, PA 15213

1 Commanding Officer
U.S. Naval Amphibious School
Coronado, CA 92155

1 Commanding Officer
Naval Health Research
Center
Attn: Library
San Diego, CA 92152

1 Naval Medical R&D Command
Code 44
National Naval Medical Center
Bethesda, MD 20014

1 Library
Navy Personnel R&D Center
San Diego, CA 92152

6 Commanding Officer
Naval Research Laboratory
Code 2627
Washington, DC 20390

1 JOHN OLSEN
CHIEF OF NAVAL EDUCATION &
TRAINING SUPPORT
PENSACOLA, FL 32509

1 Psychologist
ONR Branch Office
495 Summer Street
Boston, MA 02210

1 Psychologist
ONR Branch Office
536 S. Clark Street
Chicago, IL 60605

1 Office of Naval Research
Code 200
Arlington, VA 22217

1 Office of Naval Research
Code 441
800 N. Quincy Street
Arlington, VA 22217

5 Personnel & Training Research Programs
(Code 458)
Office of Naval Research
Arlington, VA 22217

1 Psychologist
OFFICE OF NAVAL RESEARCH BRANCH
223 OLD MARYLEBONE ROAD
LONDON, NW, 15TH ENGLAND

1 Psychologist
ONR Branch Office
1030 East Green Street
Pasadena, CA 91101

1 Scientific Director
Office of Naval Research
Scientific Liaison Group/Tokyo
American Embassy
APO San Francisco, CA 96503

1 Scientific Advisor to the Chief of
Naval Personnel (Pers-Or)
Naval Bureau of Personnel
Room 4410, Arlington Annex
Washington, DC 20370

1 LT Frank C. Petho, MSC, USNR (Ph.D)
Code L51
Naval Aerospace Medical Research Laborat
Pensacola, FL 32508

1 DR. RICHARD A. POLLAK
ACADEMIC COMPUTING CENTER
U.S. NAVAL ACADEMY
ANNAPOLIS, MD 21402

1 Roger W. Remington, Ph.D
Code L52
NAMRL
Pensacola, FL 32508

1 Dr. Bernard Rimland
Navy Personnel R&D Center
San Diego, CA 92152

1 Mr. Arnold Rubenstein
Naval Personnel Support Technology
Naval Material Command (08T244)
Room 1044, Crystal Plaza #5
2221 Jefferson Davis Highway
Arlington, VA 20360

1 Dr. Worth Scanland
Chief of Naval Education and Training
Code N-5
NAS, Pensacola, FL 32508

1 A. A. SJOHOLM
TECH. SUPPORT, CODE 201
NAVY PERSONNEL R& D CENTER
SAN DIEGO, CA 92152

1 Mr. Robert Smith
Office of Chief of Naval Operations
OP-987E
Washington, DC 20350

1 Dr. Alfred F. Smode
Training Analysis & Evaluation Group (TAEG)
Dept. of the Navy
Orlando, FL 32813

1 Dr. Richard Sorensen
Navy Personnel R&D Center
San Diego, CA 92152

1 CDR Charles J. Theisen, JR. MSC, USN
Head Human Factors Engineering Div.
Naval Air Development Center
Warminster, PA 18974

1 W. Gary Thomson
Naval Ocean Systems Center
Code 7132
San Diego, CA 92152

1 Dr. Ronald Weitzman
Department of Administrative Sciences
U. S. Naval Postgraduate School
Monterey, CA 93940

1 DR. H.M. WEST III
DEPUTY ADCNO FOR CIVILIAN PLANNING AND PROGRAMMING
RM. 2625, ARLINGTON ANNEX
WASHINGTON, DC 20370

1 DR. MARTIN F. WISKOFF
NAVY PERSONNEL R& D CENTER
SAN DIEGO, CA 92152

Army

1 HQ USAREUE & 7th Army
ODCSOPS
USAAREUE Director of GED
APO New York 09403

1 LCOL Gary Bloedorn
Training Effectiveness Analysis Division
US Army TRADOC Systems Analysis Activity
White Sands Missile Range, NM 88002

1 DR. RALPH DUSEK
U.S. ARMY RESEARCH INSTITUTE
5001 EISENHOWER AVENUE
ALEXANDRIA, VA 22333

1 Dr. Myron Fischl
U.S. Army Research Institute for the Social and Behavioral Sciences
5001 Eisenhower Avenue
Alexandria, VA 22333

1 Col. Frank Hart, Director
Training Development Institute
ATTNG-TDI
Ft. Eustis, VA 23604

1 Dr. Michael Kaplan
U.S. ARMY RESEARCH INSTITUTE
5001 EISENHOWER AVENUE
ALEXANDRIA, VA 22333

1 Dr. Milton S. Katz
Individual Training & Skill Evaluation Technical Area
U.S. Army Research Institute
5001 Eisenhower Avenue
Alexandria, VA 22333

1 Dr. Beatrice J. Farr
Army Research Institute (PERI-OK)
5001 Eisenhower Avenue
Alexandria, VA 22333

1 Dr. Milt Maier
U.S. ARMY RESEARCH INSTITUTE
5001 EISENHOWER AVENUE
ALEXANDRIA, VA 22333

1 Dr. Harold F. O'Neil, Jr.
ATTN: PERI-OK
5001 EISENHOWER AVENUE
ALEXANDRIA, VA 22333

1 LTCOL Michael T. Plummer
Organizational Effectiveness Division
Office of the Deputy Chief of Staff for Personnel
Department of the Army
Washington, DC 20301

1 Dr. Robert Sasmor
U. S. Army Research Institute for the Behavioral and Social Sciences
5001 Eisenhower Avenue
Alexandria, VA 22333

1 Dr. Frederick Steinheiser
U. S. Army Reserch Institute
5001 Eisenhower Avenue
Alexandria, VA 22333

1 Dr. Joseph Ward
U.S. Army Research Institute
5001 Eisenhower Avenue
Alexandria, VA 22333

Air Force

1 Air Force Human Resources Lab
AFHRL/PED
Brooks AFB, TX 78235

1 Air University Library
AUL/LSE 76/443
Maxwell AFB, AL 36112

1 DR. T. E. COTTERMAN
AFHRL/ASR
WRIGHT PATTERSON AFB
OHIO 45433

1 Dr. Philip De Leo
AFHRL/TT
Lowry AFB, CO 80230

1 DR. G. A. ECKSTRAND
AFHRL/AS
WRIGHT-PATTERSON AFB, OH 45433

1 Dr. Genevieve Haddad
Program Manager
Life Sciences Directorate
AFOSR
Bolling AFB, DC 20332

1 CDR. MERCER
CNET LIAISON OFFICER
AFHRL/FLYING TRAINING DIV.
WILLIAMS AFB, AZ 85224

1 Dr. Donald E. Meyer
U.S. Air Force
ATC/XPTD
Randolph AFB, TX 78148

1 Dr. Ross L. Morgan (AFHRL/ASR)
Wright -Patterson AFB
Ohio 45433

1 Dr. Roger Pennell
AFHRL/TT
Lowry AFB, CO 80230

1 Personnel Analysis Division
HQ USAF/DPXXA
Washington, DC 20330

1 Research Branch
AFMPC/DPMYP
Randolph AFB, TX 78148

1 Dr. Malcolm Ree
AFHRL/PED
Brooks AFB, TX 78235

1 Dr. Marty Rockway (AFHRL/TT)
Lowry AFB
Colorado 80230

1 Jack A. Thorpe, Capt, USAF
Program Manager
Life Sciences Directorate
AFOSR
Bolling AFB, DC 20332

1 Brian K. Waters, LCOL, USAF
Air University
Maxwell AFB
Montgomery, AL 36112

Marines

1 H. William Greenup
Education Advisor (E031)
Education Center, MCDEC
Quantico, VA 22134

1 Director, Office of Manpower Utilization
HQ, Marine Corps (MPU)
BCB, Bldg. 2009
Quantico, VA 22134

1 DR. A.L. SLAFKOSKY
SCIENTIFIC ADVISOR (CODE RD-1)
HQ, U.S. MARINE CORPS
WASHINGTON, DC 20380

CoastGuard

1 Mr. Richard Lanterman
PSYCHOLOGICAL RESEARCH (G-P-1/62)
U.S. COAST GUARD HQ
WASHINGTON, DC 20590

1 Dr. Thomas Warm
U. S. Coast Guard Institute
P. O. Substation 18
Oklahoma City, OK 73169

Other DoD

1 Dr. Stephen Andriole
ADVANCED RESEARCH PROJECTS AGENCY
1400 WILSON BLVD.
ARLINGTON, VA 22209

12 Defense Documentation Center
Cameron Station, Bldg. 5
Alexandria, VA 22314
Attn: TC

1 Dr. Dexter Fletcher
ADVANCED RESEARCH PROJECTS AGENCY
1400 WILSON BLVD.
ARLINGTON, VA 22209

1 Military Assistant for Training and
Personnel Technology
Office of the Under Secretary of Defense
for Research & Engineering
Room 3D129, The Pentagon
Washington, DC 20301

1 MAJOR Wayne Sellman, USAF
Office of the Assistant Secretary
of Defense (MRA&L)
3B930 The Pentagon
Washington, DC 20301

1 Mr. Fredrick W. Suffa
MPP (A&R)
2B269
Pentagon
Washington, D.C. 20301

Civil Govt

1 Dr. Susan Chipman
Basic Skills Program
National Institute of Education
1200 19th Street NW
Washington, DC 20208

1 Dr. Lorraine D. Eyde
Personnel R&D Center
U.S. Civil Service Commission
1900 E Street NW
Washington, D.C. 20415

11 Mr. James M. Ferstl
Bureau of Training
U.S. Civil Service Commission
Washington, D.C. 20415

1 Dr. Joseph I. Lipson
Division of Science Education
Room W-638
National Science Foundation
Washington, DC 20550

1 Dr. John Mays
National Institute of Education
1200 19th Street NW
Washington, DC 20208

1 William J. McLaurin
Rm. 301, Internal Revenue Service
2221 Jefferson Davis Highway
Arlington, VA 22202

1 Dr. Arthur Melmed
National Intitute of Education
1200 19th Street NW
Washington, DC 20208

1 Dr. Andrew R. Molnar
Science Education Dev.
and Research
National Science Foundation
Washington, DC 20550

1 Dr. Jeffrey Schiller
National Institute of Education
1200 19th St. NW
Washington, DC 20208

Civil Govt

1 Dr. Thomas G. Sticht
Basic Skills Program
National Institute of Education
1200 19th Street NW
Washington, DC 20208

1 Dr. Frank Withrow
U. S. Office of Education
400 6th Street SW
Washington, DC 20202

1 Dr. Joseph L. Young, Director
Memory & Cognitive Processes
National Science Foundation
Washington, DC 20550

Non Govt

1 Dr. Earl A. Alluisi
HQ, AFHRL (AFSC)
Brooks AFB, TX 78235

1 Dr. John R. Anderson
Department of Psychology
Carnegie Mellon University
Pittsburgh, PA 15213

1 DR. MICHAEL ATWOOD
SCIENCE APPLICATIONS INSTITUTE
40 DENVER TECH. CENTER WEST
7935 E. PRENTICE AVENUE
ENGLEWOOD, CO 80110

1 1 psychological research unit
Dept. of Defense (Army Office)
Campbell Park Offices
Canberra ACT 2600, Australia

1 Dr. R. A. Avner
University of Illinois
Computer-Based Educational Research Lab
Urbana, IL 61801

1 Dr. Alan Baddeley
Medical Research Council
Applied Psychology Unit
15 Chaucer Road
Cambridge CB2 2EF
ENGLAND

1 Dr. Patricia Baggett
Department of Psychology
University of Denver
University Park
Denver, CO 80208

1 Ms. Carole A. Bagley
Minnesota Educational Computing Consortium
2520 Broadway Drive
St. Paul, MN 55113

1 Dr. John Bergan
School of Education
University of Arizona
Tuscon AZ 85721

1 Dr. Nicholas A. Bond
Dept. of Psychology
Sacramento State College
600 Jay Street
Sacramento, CA 95819

1 Dr. Lyle Bourne
Department of Psychology
University of Colorado
Boulder, CO 80302

1 Dr. Kenneth Bowles
Institute for Information Sciences
University of California at San Diego
La Jolla, CA 92037

1 Dr. John Brackett
SofTech
460 Totten Pond Road
Waltham, MA 02154

1 Dr. John S. Brown
XEROX Palo Alto Research Center
3333 Coyote Road
Palo Alto, CA 94304

1 DR. C. VICTOR BUNDERSON
WICAT INC.
UNIVERSITY PLAZA, SUITE 10
1160 SO. STATE ST.
OREM, UT 84057

1 Dr. Anthony Cancelli
School of Education
University of Arizona
Tuscon, AZ 85721

1 Dr. John B. Carroll
Psychometric Lab
Univ. of No. Carolina
Davie Hall 013A
Chapel Hill, NC 27514

1 Charles Myers Library
Livingstone House
Livingstone Road
Stratford
London E15 2LJ
ENGLAND

1 Dr. William Chase
Department of Psychology
Carnegie Mellon University
Pittsburgh, PA 15213

1 Dr. Micheline Chi
Learning R & D Center
University of Pittsburgh
3939 O'Hara Street
Pittsburgh, PA 15213

1 Dr. John Chiorini
Litton-Mellonics
Box 1286
Springfield, VA 22151

1 Dr. William Clancey
Department of Computer Science
Stanford University
Stanford, CA 94305

1 Dr. Kenneth E. Clark
College of Arts & Sciences
University of Rochester
River Campus Station
Rochester, NY 14627

1 Dr. Norman Cliff
Dept. of Psychology
Univ. of So. California
University Park
Los Angeles, CA 90007

1 Dr. Allan M. Collins
Bolt Beranek & Newman, Inc.
50 Moulton Street
Cambridge, Ma 02138

1 Dr. Meredith Crawford
Department of Engineering Administration
George Washington University
Suite 805
2101 L Street N. W.
Washington, DC 20037

1 Dr. Ruth Day
Center for Advanced Study
in Behavioral Sciences
202 Junipero Serra Blvd.
Stanford, CA 94305

1 Dr. Hubert Dreyfus
Department of Philosophy
University of California
Berkely, CA 94720

1 Dr. Marvin D. Dunnette
N492 Elliott Hall
Dept. of Psychology
Univ. of Minnesota
Minneapolis, MN 55455

1 Dr. A. J. Eschenbrenner
Dept. E422, Bldg. 101
McDonnell Douglas Astronautics Co.
P.O.Box 516
St. Louis, MO 63166

1 MAJOR I. N. EVONIC
CANADIAN FORCES PERS. APPLIED RESEARCH
1107 AVENUE ROAD
TORONTO, ONTARIO, CANADA

1 Dr. Leonard Feldt
Lindquist Center for Measurment
University of Iowa
Iowa City, IA 52242

1 Dr. Richard L. Ferguson
The American College Testing Program
P.O. Box 168
Iowa City, IA 52240

1 Mr. Wallace Feurzeig
Bolt Beranek & Newman, Inc.
50 Moulton St.
Cambridge, MA 02138

1 Dr. Victor Fields
Dept. of Psychology
Montgomery College
Rockville, MD 20850

1 Dr. Edwin A. Fleishman
Advanced Research Resources Organ.
Suite 900
4330 East West Highway
Washington, DC 20014

1 Dr. John R. Frederiksen
Bolt Beranek & Newman
50 Moulton Street
Cambridge, MA 02138

1 Dr. Alinda Friedman
Department of Psychology
University of Alberta
Edmonton, Alberta
CANADA T6G 2J9

1 Dr. Vernon S. Gerlach
College of Education
146 Payne Bldg. B
Arizona State University
Tempe, AZ 85281

1 DR. ROBERT GLASER
LRDC
UNIVERSITY OF PITTSBURGH
3939 O'HARA STREET
PITTSBURGH, PA 15213

1 DR. JAMES G. GREENO
LRDC
UNIVERSITY OF PITTSBURGH
3939 O'HARA STREET
PITTSBURGH, PA 15213

1 Dr. Ron Hambleton
School of Education
University of Massechusetts
Amherst, MA 01002

1 Dr. Harold Hawkins
Department of Psychology
University of Oregon
Eugene OR 97403

1 Dr. Dustin H. Heuston
Wicat, Inc.
Box 986
Orem, UT 84057

1 Dr. Lloyd Humphreys
Department of Psychology
University of Illinois
Champaign, IL 61820

1. Library
HumRRO/Western Division
27857 Berwick Drive
Carmel, CA 93921

1 Dr. Earl Hunt
Dept. of Psychology
University of Washington
Seattle, WA 98105

1 DR. LAWRENCE B. JOHNSON
LAWRENCE JOHNSON & ASSOC., INC.
SUITE 502
2001 S STREET NW
WASHINGTON, DC 20009

1 Dr. Wilson A. Judd
McDonnell-Douglas
Astronautics Co. East
Lowry AFB
Denver, CO 80230

1 Dr. Arnold F. Kanarick
Honeywell, Inc.
2600 Ridgeway Pkwy
Minneapolis, MN 55413

1 Dr. Steven W. Keele
Dept. of Psychology
University of Oregon
Eugene, OR 97403

1 Dr. Walter Kintsch
Department of Psychology
University of Colorado
Boulder, CO 80302

1 Dr. David Kieras
Department of Psychology
University of Arizona
Tuscon, AZ 85721

1 Dr. Stephen Kosslyn
Harvard University
Department of Psychology
33 Kirkland Street
Cambridge, MA 02138

1 Mr. Marlin Kroger
1117 Via Goleta
Palos Verdes Estates, CA 90274

1 LCOL. C.R.J. LAFLEUR
PERSONNEL APPLIED RESEARCH
NATIONAL DEFENSE HQS
101 COLONEL BY DRIVE
OTTAWA, CANADA K1A 0K2

1 Dr. Jill Larkin
Department of Psychology
Carnegie Mellon University
Pittsburgh, PA 15213

1 Dr. Alan Lesgold
Learning R&D Center
University of Pittsburgh
Pittsburgh, PA 15260

1 Dr. Robert Linn
College of Education
University of Illinois
Urbana, IL 61801

1 Dr. Robert R. Mackie
Human Factors Research, Inc.
6780 Cortona Drive
Santa Barbara Research Pk.
Goleta, CA 93017

1 Dr. Richard B. Millward
Dept. of Psychology
Hunter Lab.
Brown University
Providence, RI 82912

1 Dr. Stuart Milner
Department of Education
George Mason University
4400 Fairfax Drive
Fairfax, VA 22030

1 Dr. Allen Munro
Univ. of So. California
Behavioral Technology Labs
3717 South Hope Street
Los Angeles, CA 90007

1 Dr. Donald A Norman
Dept. of Psychology C-009
Univ. of California, San Diego
La Jolla, CA 92093

1 Dr. Melvin R. Novick
Iowa Testing Programs
University of Iowa
Iowa City, IA 52242

1 Dr. Jesse Orlansky
Institute for Defense Analysis
400 Army Navy Drive
Arlington, VA 22202

1 Dr. Robert Pachella
Department of Psychology
Human Performance Center
330 Packard Road
Ann Arbor, MI 48104

1 Dr. James A. Paulson
Portland State University
P.O. Box 751
Portland, OR 97207

1 MR. LUIGI PETRULLO
2431 N. EDGEWOOD STREET
ARLINGTON, VA 22207

1 Dr. Barbara Pflanz
Department of German
University of Redlands
Redlands, CA 92373

1 Dr. Martha Polson
Department of Psychology
University of Colorado
Boulder, CO 80302

1 DR. PETER POLSON
DEPT. OF PSYCHOLOGY
UNIVERSITY OF COLORADO
BOULDER, CO 80302

1 DR. DIANE M. RAMSEY-KLEE
R-K RESEARCH & SYSTEM DESIGN
3947 RIDGEMONT DRIVE
MALIBU, CA 90265

1 MIN. RET. M. RAUCH
P II 4
BUNDESMINISTERIUM DER VERTEIDIGUNG
POSTFACH 161
53 BONN 1, GERMANY

1 Dr. Peter B. Read
Social Science Research Council
605 Third Avenue
New York, NY 10016

1 Dr. Mark D. Reckase
Educational Psychology Dept.
University of Missouri-Columbia
12 Hill Hall
Columbia, MO 65201

1 Dr. Fred Reif
SESAME
c/o Physics Department
University of California
Berkely, CA 94720

1 Dr. Andrew M. Rose
American Institutes for Research
1055 Thomas Jefferson St. NW
Washington, DC 20007

1 Dr. Ernst Z. Rothkopf
Bell Laboratories
600 Mountain Avenue
Murray Hill, NJ 07974

1 Dr. David Rumelhart
Center for Human Information Processing
Univ. of California, San Diego
La Jolla, CA 92093

1 PROF. FUMIKO SAMEJIMA
DEPT. OF PSYCHOLOGY
UNIVERSITY OF TENNESSEE
KNOXVILLE, TN 37916

1 Dr. Benjamin Schneider
Dept. of Psychology
Univ. of Maryland
College Park, MD 20742

1 DR. WALTER SCHNEIDER
DEPT. OF PSYCHOLOGY
UNIVERSITY OF ILLINOIS
CHAMPAIGN, IL 61820

1 DR. ROBERT J. SEIDEL
INSTRUCTIONAL TECHNOLOGY GROUP
HUMRRO
300 N. WASHINGTON ST.
ALEXANDRIA, VA 22314

1 Dr. Robert Smith
Department of Computer Science
Rutgers University
New Brunswick, NJ 08903

1 Dr. Robert Sternberg
Dept. of Psychology
Yale University
Box 11A, Yale Station
New Haven, CT 06520

1 DR. ALBERT STEVENS
BOLT BERANEK & NEWMAN, INC.
50 MOULTON STREET
CAMBRIDGE, MA 02138

1 Mr. William Stobie
McDonnell-Douglas
Astronautics Co.
P. O. Box 30204
Chico, CA 95926

1 DR. PATRICK SUPPES
INSTITUTE FOR MATHEMATICAL STUDIES IN
THE SOCIAL SCIENCES
STANFORD UNIVERSITY
STANFORD, CA 94305

1 Dr. Hariharan Swaminathan
Laboratory of Psychometric and
Evaluation Research
School of Education
University of Massachusetts
Amherst, MA 01003

1 Dr. Brad Sympson
Office of Data Analysis Research
Educational Testing Service
Princeton, NJ 08541

1 Dr. Kikumi Tatsuoka
Computer Based Education Research
Laboratory
252 Engineering Research Laboratory
University of Illinois
Urbana, IL 61801

1 Dr. Maurice Tatsuoka
Department of Educational Psychology
University of Illinois
Champaign, IL 61801

1 Dr. David Thissen
Department of Psychology
University of Kansas
Lawrence, KS 66044

1 DR. PERRY THORNDYKE
THE RAND CORPORATION
1700 MAIN STREET
SANTA MONICA, CA 90406